

WO9504069

Publication Title:

BIOTINYLATION OF PROTEINS

Abstract:

Biotinylation peptides can be fused to other peptides or proteins of interest using recombinant DNA techniques to provide efficient methods for biotinylating the resulting fusion proteins in vivo or in vitro.

Data supplied from the esp@cenet database - <http://ep.espacenet.com>

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁵ : C07H 21/04, C12N 15/11, 15/63, 15/65	A1	(11) International Publication Number: WO 95/04069 (43) International Publication Date: 9 February 1995 (09.02.95)
(21) International Application Number: PCT/US94/08528 (22) International Filing Date: 28 July 1994 (28.07.94) (30) Priority Data: 08/099,991 30 July 1993 (30.07.93) US (60) Parent Application or Grant (63) Related by Continuation US 08/099,991 (CIP) Filed on 30 July 1993 (30.07.93) (71) Applicant (for all designated States except US): AFFYMAX TECHNOLOGIES N.V. [NL/NL]; De Ruyderkade 62, Cu- racao (AN). (72) Inventor; and (75) Inventor/Applicant (for US only): SCHATZ, Peter, J. [US/US]; 2080 Marich Way #15, Mountain View, CA 94040 (US). (74) Agents: NORVIEL, Vern; Townsend and Townsend Kourie and Crew, 20th floor, One Market Plaza, Steuart Street Tower, San Francisco, CA 94105 (US) et al.		(81) Designated States: AU, CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report.</i>
(54) Title: BIOTINYLATION OF PROTEINS (57) Abstract Biotinylation peptides can be fused to other peptides or proteins of interest using recombinant DNA techniques to provide efficient methods for biotinylating the resulting fusion proteins <i>in vivo</i> or <i>in vitro</i> .		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

BIOTINYLATION OF PROTEINS

5

BACKGROUND OF THE INVENTION

CROSS-REFERENCE TO RELATED APPLICATION

This application is a continuation-in-part
10 application of U.S. Serial Number 08/099,991, filed July 30,
1993, which is hereby incorporated by reference and benefit is
claimed of its filing date.

1. Field of the Invention

The present invention relates to methods for
15 producing biotinylated proteins in vitro and in recombinant
host cells. The invention therefore relates to the field of
molecular biology, but given the diverse uses for recombinant
proteins, the invention also relates to the fields of
chemistry, pharmacology, biotechnology, and medical
20 diagnostics.

2. Description of the Background Art

The ability to synthesize DNA chemically has made
possible the construction of peptides and proteins not
25 otherwise found in nature and useful in a wide variety of
methods that would otherwise be very difficult or impossible
to perform. One illustrative example of this technology
relates to the class of molecules known as receptors.
Receptor proteins mediate important biological functions
30 through interactions with ligands. For many years,
researchers have attempted to isolate and identify ligands
that interact with receptors in ways that can help ameliorate
human (and other) disease. The advent of molecular biology
has revolutionized the way these researchers study
35 receptor-ligand interaction. For instance, standard molecular
biology techniques have enabled the cloning and high-level
expression of many receptors in recombinant host cells.

The patent literature, for instance, is replete with publications describing the recombinant expression of receptor proteins. See, e.g., PCT Patent Pub. No. 91/18982 and U.S. Patent Nos. 5,081,228 and 4,968,607, which describe
5 recombinant DNA molecules encoding the IL-1 receptor; U.S. Patent Nos. 4,816,565; 4,578,335; and 4,845,198, which describe recombinant DNA and proteins relating to the IL-2 receptor; PCT Patent Pub. No. 91/08214, which describes EGF receptor gene related nucleic acids; PCT Patent Pub. No.
10 91/16431 and U.S. Patent No. 4,897,264, which describe the interferon gamma receptor and related proteins and nucleic acids; European Patent Office (EPO) Pub. No. 377,489, which describes the C5a receptor protein; PCT Patent Pub. No. 90/08822, which describes the EPO receptor and related nucleic
15 acids; and PCT Patent Pub. No. 92/01715, which describes MHC receptors.

Several of the above publications not only describe how to isolate a particular receptor protein (or the gene encoding the protein) but also describe variants of the
20 receptor that may be useful in ways the natural or native receptor is not. For instance, PCT Patent Pub. No. 91/16431 describes soluble versions of the gamma interferon receptor, while PCT Patent Pub. No. 92/01715 describes how to produce soluble cell-surface dimeric proteins. This latter technology
25 involves expression of the receptor with a signal for lipid attachment; once the lipid is attached to the receptor, the receptor becomes anchored in the cell membrane, where the dimeric form of the receptor is assembled. See also U.S. Serial No. 947,339, filed on September 18, 1992, and
30 incorporated herein by reference for all purposes, which describes how HPAP-containing receptors can be cleaved from the cell surface and how the anchoring sequences that remain can serve as recognition sequences for antibodies that are used to immobilize the receptor.

35 The advances made with respect to receptor cloning and expression have been accompanied by advances in technology relating to methods for screening a receptor against compounds that may interact with the receptor in a desired fashion. One

such advance relates to the generation of large numbers of compounds, or potential ligands, in a variety of random and semi-random "peptide diversity" generation systems. These systems include the "peptides on plasmids" system described in U.S. patent application Serial No. 963,321, filed October 15, 1992, which is a continuation-in-part of U.S. patent application Serial No. 778,233, filed October 16, 1991; the "peptides on phage" system described in U.S. Patent Application Serial No. 718,577, filed June 20, 1991, which is a continuation-in-part of Serial No. 541,108, filed June 20, 1990; Cwirla *et al.*, August 1990, Proc. Natl. Acad. Sci. USA 87: 6378-6382; Barrett *et al.*, 1992, Analyt. Biochem. 204: 357-364; and PCT Patent Pub. Nos. 91/18980 and 91/19818; the phage-based antibody display systems described in U.S. patent application Serial No. 517,659, filed May 11, 1990, and PCT Patent Pub. No. 91/17271; the bead-based systems for generating and screening nucleic acid ligands described in PCT Pub. Nos. 91/19813, 92/05258, and 92/14843; the bead-based system described in U.S. Patent Application Serial No. 946,239, filed September 16, 1992, which is a continuation-in-part of Serial No. 762,522, filed September 18, 1991; and the "very large scaled immobilized polymer synthesis" system described in U.S. Patent No. 5,143,854; PCT Patent Pub. Nos. 90/15070 and 92/10092; U.S. Patent Application Serial No. 624,120, filed December 6, 1990; Fodor *et al.*, 15 February 1991, Science 251: 767-773; Dower and Fodor, 1991, Ann. Rep. Med. Chem. 26:271-180; and U.S. patent application Serial No. 805,727, filed December 6, 1991. Each of the above references is incorporated herein by reference for all purposes.

Other developments relate to how the receptor is used in such screening methods. One important advance relates to the development of reagents and methods for immobilizing one or more receptors in a spatially defined array, as described in PCT Patent Pub. No. 91/07087. In one embodiment of this method, a receptor is attached to avidin and then immobilized on a surface that bears biotin groups. The surface is first prepared, however, with caged biotin groups,

which will not bind avidin until the caging group is removed by, in this embodiment, irradiation. Once the avidinylated receptor is bound to the biotin groups on the surface, the surface can be used in screening compounds against the receptor.

Biotin is a coenzyme that is covalently attached to several enzymes involved in the transfer of activated carboxyl groups. As the above example illustrates, biotin labeling of molecules not normally biotinylated can be used to label, detect, purify, and/or immobilize such molecules. These methods also rely upon the proteins avidin and streptavidin, which bind very tightly and specifically to biotin and other biotin-binding molecules, some of which bind to biotin with different affinity than avidin. Typically, the biotinylated molecules used in such methods are prepared by an in vitro biotinylation process. A method for biotinylating proteins synthesized by recombinant DNA techniques in vivo would eliminate the need to biotinylate these proteins chemically after purification and would greatly simplify the purification process, due to the ability to use the biotin as an affinity tag (see Green, 1975, Adv. Protein Res. 29:85-133, incorporated herein by reference).

Biotin is added to proteins in vivo through the formation of an amide bond between the biotin carboxyl group and the epsilon-amino group of specific lysine residues in a reaction that requires ATP. In normal E. coli, only one protein is biotinylated, the biotin carboxyl carrier protein (BCCP) subunit of acetyl-CoA carboxylase. This reaction is catalyzed by the biotin-protein ligase (BirA), the product of the birA gene (see Cronan, 1989, Cell 58: 427-429, incorporated herein by reference).

Others have proposed a means by which biotin labeling can be accomplished in vivo by the addition of a domain of at least 75 amino acids to recombinant proteins (see Cronan, 1990, J. Biol. Chem. 265: 10327-10333, incorporated herein by reference). See also Cress et al., 1993, Promega Notes 42: 2-7. Addition of this 75 amino acid domain to several different proteins leads to the biotinylation of the

fusion proteins by BirA on a specific lysine of the added domain. Addition of smaller fragments of the 75 residue domain does not lead to biotinylation, implying that a reasonably complex recognition domain is required. Changes in the sequence of biotinylated proteins as far as 33 residues from the modified lysine abolish biotinylation (see Murtif and Samols, 1987, *J. Biol. Chem.* 262: 11813-11816). Changes close to the lysine also affect biotinylation (see Shenoy *et al.*, 1988, *FASEB J.* 2: 2505-2511, and Shenoy *et al.*, 1992, *J. Biol. Chem.* 267: 18407-18412). Unfortunately, however, the addition of such a large protein domain may negatively affect the biochemical properties of a biotinylated protein. Smaller domains that specify biotinylation would be very beneficial, in that such domains would have a minimal structural effect on the wide variety of possible fusion partners. Also, the 75 residue domain does not lead to complete biotinylation of the domain, and improved domains could be more efficient. The present invention provides such improved biotinylation domains.

SUMMARY OF THE INVENTION

The present invention provides useful compounds, reagents, methods, and kits for biotinylating proteins. In a first aspect, the present invention provides a method for biotinylating a protein, said method comprising: (a) constructing a recombinant DNA expression vector that encodes a fusion protein comprising said protein and a biotinylation peptide less than 50 amino acids in length; (b) transforming a recombinant host cell capable of synthesizing a biotinylation enzyme with said vector; and (c) culturing said host cell under conditions in which biotin is present and such that said fusion protein and biotinylation enzyme are expressed, resulting in biotinylation of said fusion protein. If the host cell does not naturally produce biotin, then one can add biotin to the media. In a preferred embodiment, the host cell is *E. coli*, and the biotinylation enzyme is BirA.

Thus, in the preferred embodiment, a biotinylation peptide of the present invention can be added to any protein

expressed in E. coli with a sufficient time of retention in the cytoplasm to permit BirA to act. If high expression levels of biotinylated protein are desired, then one can readily overexpress the BirA protein at the same time (see 5 Buoncristiani et al., 1988, J. Biol. Chem. 263, 1013-1016, incorporated herein by reference). In similar fashion, host cells that lack an endogenous biotin protein ligase (called a biotinylation enzyme) can be transformed with a vector that codes for expression of the birA gene to provide or enhance 10 their ability to biotinylate recombinant proteins. Where, due to the conservation of the recognition domains, the endogenous biotin-protein ligase of other non-E. coli cell types recognize the novel biotinylation sequences, no such recombinant expression of a biotinylation enzyme is required. 15 One can also perform the biotinylation reaction in vitro using a biotinylation enzyme such as purified BirA (see Buoncristiani, supra), biotin, and biotinylation sequence peptide-tagged proteins, which proteins may be either produced in recombinant host cells or by in vitro translation. One can 20 also use biotin analogues, such as 2-iminobiotin, which has a lower affinity for avidin than biotin and so may be preferred for some applications, in place of biotin, in the method.

The present invention also provides reagents useful in the present method, including peptides, proteins, 25 oligonucleotides, and recombinant DNA expression vectors. Thus, the present invention provides biotinylated peptides less than 50 amino acids in length, typically 10 to 20 or more amino acids in length, and oligonucleotides comprising coding sequences for such peptides. In addition, the invention 30 provides recombinant biotinylated proteins and expression vectors encoding those proteins. In a preferred embodiment the present biotinylation peptide is 13 amino acids long and is defined by $LX_1X_2IX_3X_4X_5X_6KX_7X_8X_9X_{10}$ (SEQ. ID NO:1), where X_1 is any amino acid, X_2 is any amino acid other than large 35 hydrophobic amino acids (such as L, V, I, W, F, Y); X_3 is F or L, X_4 is E or D; X_5 is A, G, S, or T; X_6 is Q or M; X_7 is I, M, or V; X_8 is E, L, V, Y, or I; X_9 is W, Y, V, F, L, or I;

and X₁₀ is preferably R or H but may be any amino acid other than acidic amino acids such as D or E.

In summary, this invention provides a simple and efficient means to biotinylate recombinant proteins, providing for rapid purification, immobilization, labeling, and detection of those proteins. The method is useful for a variety of purposes and is widely commercially useful for research and diagnostic applications.

DESCRIPTION OF THE PREFERRED EMBODIMENT

I. Definitions

For purposes of understanding the present invention, the following terms are defined.

Amino acid residues in peptides are abbreviated as follows: Phenylalanine is Phe or F; Leucine is Leu or L; Isoleucine is Ile or I; Methionine is Met or M; Valine is Val or V; Serine is Ser or S; Proline is Pro or P; Threonine is Thr or T; Alanine is Ala or A; Tyrosine is Tyr or Y; Histidine is His or H; Glutamine is Gln or Q; Asparagine is Asn or N; Lysine is Lys or K; Aspartic Acid is Asp or D; Glutamic Acid is Glu or E; Cysteine is Cys or C; Tryptophan is Trp or W; Arginine is Arg or R; and Glycine is Gly or G.

The term "antibody" refers to antibodies and antibody fragments that retain the ability to bind the epitope that the intact antibody binds, whether the antibody or fragment is produced by hybridoma cell lines, by immunization to elicit a polyclonal antibody response, or by recombinant host cells that have been transformed with a recombinant DNA expression vector that encodes the antibody or antibody fragment.

The term "antigen" is defined as a molecule that induces the formation of an antibody or is capable of binding specifically to the antigen-binding sites of an antibody.

The term "effective amount" refers to an amount sufficient to induce a desired result.

The term "epitope" refers to that portion of an antigen that interacts with an antibody.

The term "host cell" refers to a eukaryotic or procaryotic cell or group of cells that can be or has been transformed by a recombinant DNA vector. For purposes of the present invention, procaryotic host cells are preferred.

5 The term "ligand" refers to a molecule that is recognized by a particular receptor. The agent bound by or reacting with a receptor is called a "ligand," a term which is definitionally meaningful primarily in terms of its counterpart receptor. The term "ligand" does not imply any particular molecular size or other structural or compositional
10 feature other than that the substance in question is capable of binding or otherwise interacting with the receptor. A "ligand" may serve either as the natural ligand to which the receptor binds or as a functional analogue that may act as an agonist or antagonist. Examples of ligands that can be
15 investigated with the present invention include, but are not restricted to, peptides and proteins such as agonists and antagonists for cell membrane receptors, toxins and venoms, epitopes such as viral epitopes, antibodies, hormones, enzyme substrates, and proteins.

20 The term "linker" or "spacer" refers to a molecule or group of molecules (such as a monomer or polymer) that connects two molecules and often serves to place the two molecules in a preferred configuration, e.g., so that a ligand
25 can bind to a receptor with minimal steric hindrance.

The term "monomer" refers to any member of the set of molecules that can be joined together to form an oligomer or polymer. The set of monomers useful in the present invention includes, but is not restricted to, for the example
30 of peptide synthesis, the set of L-amino acids, D-amino acids, or synthetic amino acids. As used herein, "monomer" refers to any member of a basis set for synthesis of an oligomer. For example, dimers of L-amino acids form a basis set of 400
35 "monomers" for synthesis of polypeptides. Different basis sets of monomers may be used at successive steps in the synthesis of a polymer. The term "monomer" also refers to a chemical subunit that can be combined with a different

chemical subunit to form a compound larger than either subunit alone.

The term "oligomer" or "polymer" refers to the compounds formed by the chemical or enzymatic addition of two or more monomers to one another. Such oligomers include, for example, both linear, cyclic, and branched polymers of nucleic acids and peptides, which peptides can have either alpha-, beta-, or omega-amino acids.

The term "oligonucleotide" refers to a single-stranded DNA or RNA molecule or to analogs of either. Suitable oligonucleotides may be prepared by the phosphoramidite method described by Beaucage *et al.*, 1981, *Tetr. Lett.* 22:1859-1862, or by the triester method, according to Matteucci *et al.*, 1981, *J. Am. Chem. Soc.* 103:3185, or by other methods, such as by using commercially available, automated oligonucleotide synthesizers.

The term "operably linked" refers to the placement of one nucleic acid into a functional relationship with another nucleic acid. For instance, a promoter is "operably linked" to a coding sequence if the promoter causes the transcription of the coding sequence. Generally, "operably linked" means that the DNA sequences being linked are contiguous and, where necessary to join two peptide or protein coding regions, in reading frame with one another.

The term "peptide" refers to an oligomer in which the monomers are amino acids (usually alpha-amino acids) joined together through amide bonds. Alternatively, a "peptide" can be referred to as a "polypeptide." Peptides are more than two amino acid monomers long, but more often are more than 5 to 10 amino acid monomers long and can be even longer than 20 amino acids, although peptides longer than 20 amino acids are more likely to be called "polypeptides."

The term "protein" is well known in the art and usually refers to a very large polypeptide, or set of associated polypeptides, that has some biological function. For purposes of the present invention the terms "peptide," "polypeptide," and "protein" are largely interchangeable as

libraries of all three types can be prepared using substantially similar methodology.

The term "random peptide" refers to an oligomer composed of two or more amino acid monomers and constructed by a means with which one does not entirely preselect the specific sequence of any particular oligomer. The term "random peptide library" refers not only to a set of recombinant DNA vectors that encodes a set of random peptides, but also to the set of random peptides encoded by those vectors, as well as the fusion proteins containing those random peptides. The term "protein library" has a meaning similar to "random peptide library," but the different library members differ with respect to the amino acid sequence of, or coding sequence for, the protein of interest, so that the library serves as a collection of related but different versions of the same protein.

The term "receptor" refers to a molecule that has an affinity for a given ligand. Receptors may be naturally-occurring or synthetic molecules. Receptors can be employed in their unaltered natural or isolated state, in a recombinant or modified form, or as aggregates with other species. Examples of receptors that can be employed in the method of the present invention include, but are not restricted to, antibodies, cell membrane receptors, monoclonal antibodies, antisera reactive with specific antigenic determinants (such as on viruses, cells, or other materials), polynucleotides, nucleic acids, lectins, polysaccharides, cells, cellular membranes, viruses, and organelles. Receptors are sometimes referred to in the art as "anti-ligands." As the term "receptor" is used herein, no difference in meaning is intended. A "ligand-receptor pair" is formed when a receptor and ligand have combined through molecular recognition to form a complex.

The terms "recombinant DNA cloning vector" and "recombinant DNA expression vector" refer to a DNA or RNA molecule that encodes a useful function and can either be used to transform a host cell or be introduced into a cell-free translation system to produce a protein encoded by the vector.

For purposes of the present invention, a cloning vector typically serves primarily as an intermediate in the construction of an expression vector; the latter vector is used to transform or transfect a host cell (or is introduced into a cell-free transcription and translation system) so that the transformed host cell (or cell-free transcription and translation system) produces a protein or other product encoded by the vector. Such vectors are typically "plasmids," which, for purposes of the present invention, are vectors that can be extrachromosomally maintained in a host cell, but can also be vectors that integrate into the genome of a host cell. Those of skill in the art may refer to "cloning vectors", as defined herein, as "vectors" and to "expression vectors," as defined herein, as "plasmids."

The term "solid support" refers to a material having a rigid or semi-rigid surface. Such materials will preferably take the form of small beads, pellets, disks, chips, or wafers, although other forms may be used. In some embodiments, at least one surface of the solid support will be substantially flat.

The term "surface" refers to any generally two-dimensional structure on a solid substrate and may have steps, ridges, kinks, terraces, and the like without ceasing to be a surface.

The term "synthetic" refers to production by in vitro chemical or enzymatic synthesis.

II. Methods and Reagents of the Invention

The random peptide generating and screening system known as the "peptides on plasmids" system was used to discover the small, efficient peptide biotinylation sequences of the present invention. The library was constructed to express peptides of the form: $X_{10}\underline{IVXAMKMX}_{10}$ (SEQ. ID NO:2), where X denotes a random residue, the other letters are single-letter abbreviations of amino acids, and the underlining indicates slight degeneracy in the codon for the specified amino acids, as described below. This sequence was selected based on the known sequences of several biotinylated

proteins (see, Samols et al., 1988, J. Biol. Chem. 263:6461-6464, incorporated herein by reference) as shown in Table 1. As denoted by the ellipses, the sequences below are only portions of the large sequences believed, prior to the present invention, to be necessary for biotinylation.

Table 1

TC 1.3S	...GQTVLVLEAMKMETEINAPTDG...	(SEQ. ID NO:3)
OADC	...GEVLLILEAMKMETEIRAAQAG...	(SEQ. ID NO:4)
cACC	...GQCFAEIEVMKMVMTLTAGESG...	(SEQ. ID NO:5)
EcBCCP	...GNTLCIVEAMKMMNQIEADKSG...	(SEQ. ID NO:6)
yPC	...GQPVAVL SAMKMEMI ISSPSDG...	(SEQ. ID NO:7)
hPC	...GQPLCVLSAMKMETVVTSPMEG...	(SEQ. ID NO:8)
sPC	... GQPLVLSAMKMETVVTSPVTE...	(SEQ. ID NO:9)
aPC	... GAPLVLSAMKMETVVTAPR ...	(SEQ. ID NO:10)
hPCC	...GQEICVIEAMKMQNSMTAGKTG...	(SEQ. ID NO:11)
tbp	...GQPVLVLEAMKMEHVVKAPANG...	(SEQ. ID NO:12)

The lysine residue that becomes biotinylated is contained within the "AMKM" (SEQ. ID NO:13) sequence common to most of the proteins in Table 1, which are the 1.3S subunit of Propionibacterium shermanii transcarboxylase (TC 1.3S); the Klebsiella oxaloacetate decarboxylase (OADC); chicken acetyl-CoA carboxylase (cACC); the E. coli acetyl-CoA carboxylase (EcBCCP); the yeast pyruvate carboxylase (yPC); the human pyruvate carboxylase (hPC); the sheep pyruvate carboxylase (sPC); the rat pyruvate carboxylase (aPC); the human propionyl-CoA carboxylase (hPCC); and the tomato biotinyl peptide (tbp). The sequences of these proteins share several conserved residues and/or regions having similar properties (e.g., branched chain amino acids or amidated acids).

Despite this teaching that a large region is required for biotinylation, the peptides on plasmids library used to discover the biotinylation peptides of the present invention was designed to display random peptides only 27 amino acids long, containing only one fixed codon (for K) and 5 conserved codons (for the underlined amino acids above).

Conserved codons were prepared by programming the oligonucleotide synthesizer to add, for each nucleotide of a conserved codon, 91% of the correct nucleotide and 3% each of the three other nucleotides. By this method, a very large library of random peptides was prepared and used to transform *E. coli* host cells. The peptides encoded by each clone of these libraries were fused to the carboxy-terminus of the sequence-specific DNA binding protein lac repressor (LacI). Each library particle consists of a LacI-peptide fusion bound to the lac operator (lacO) sites on the same plasmid that encoded it. Expression of these libraries in the cytoplasm allows the cells to provide compartmentalization, so that each fusion protein is bound to the appropriate plasmid.

Because the peptides on plasmids library particles are cytoplasmic, the random peptide region has access to the BirA enzyme in *E. coli* host cells. Any random peptides that productively interact with BirA (presumably a small fraction of the total) become biotinylated. After cell lysis, the biotinylated library particles were isolated by binding to immobilized streptavidin. The background of peptide sequences that bind to streptavidin in the absence of biotinylation (see Devlin *et al.*, 1990, Science 249: 404-406, incorporated herein by reference) were eliminated by adding free biotin competitor after allowing the library particles to bind to the immobilized streptavidin. The affinity of these background peptides for streptavidin is lower than the affinity of biotin, and so background peptides are displaced by the free biotin. The desired biotinylated peptides were not displaced by biotin, because those peptides are allowed to bind first, have an affinity similar to that of biotin, and interact multivalently with the immobilized streptavidin.

Thus, the protocol involved lysing the transformed cells, removing cellular debris by centrifugation, and collecting the crude lysate, from which plasmids encoding biotinylation peptides were isolated by affinity enrichment on streptavidin, as described in the Examples below. This process was repeated three times, starting with the plasmids

collected at the end of each previous cycle, with the results shown in Table 2.

Table 2

	<u>Cycle 1</u>	<u>Cycle 2</u>	<u>Cycle 3</u>	<u>Cycle 4</u>
Input plasmids	6.8×10^{11}	2.3×10^{11}	1.65×10^{11}	8.5×10^9
Recovered plasmids	7.6×10^5	1.4×10^6	3.0×10^6	1.42×10^7
% Recovered	0.00011	0.006	0.0018	0.166
Negative Control Recovered Plasmids	N.A.	3.8×10^5	3.8×10^5	1.1×10^6
% Recovered (Negative Control)	N.A.	0.00017	0.00022	0.013
Enrichment Factor	N.A.	3.6	8	13

These results indicated that the library contained members that displayed biotinylation peptides and so could be enriched and identified. Several of the isolates from the fourth round of streptavidin binding were tested to determine whether the displayed peptides directed biotinylation. The sequences of the random peptides in the positive clones are shown in Table 3, ranked in order of the strength of their reaction in an ELISA. The sequences reveal not only residues that tend to move closer to the consensus sequence defined by the known biotinylated proteins but also residues that are different from the known consensus sequence. The peptides do not have sequence motifs (such as HPQ) that have been associated with weak binding to streptavidin (Devlin *et al.*, *supra*).

Table 3

	LEEVDSTSSAIFDAMKMVWISPTEFR	(SEQ. ID NO:14)
	QGDRDETLPMILRAMKMEVYNPGGHEK	(SEQ. ID NO:15)
	SKCSYSHDLKIFEAQKMLVHSYLRVMYNY	(SEQ. ID NO:16)
5	MASSDDGLLTIFDATKMMFIRT	(SEQ. ID NO:17)
	SYMDRTDVPTILEAMKMEHTTPWACR	(SEQ. ID NO:18)
	SFPPSLPDKNIFEAMKMYVIT	(SEQ. ID NO:19)
	SVVPEPGWDGPFESMKMVYHSGAQSGQ	(SEQ. ID NO:20)
	VRHLPPPLPALFDAMKMEFVTSVQF	(SEQ. ID NO:21)
10	DMTMPTGMTKIFEAMKMEVST	(SEQ. ID NO:22)
	ATAGPLHEPDIFLAMKMEVVDVTNKAGQ	(SEQ. ID NO:23)
	SMWETLNAQKTVLL	(SEQ. ID NO:24)
	SHPSQLMTNDIFEGMKMLYH	(SEQ. ID NO:25)
	SIERGGSTHKILAAMKMYQVSTPSCS	(SEQ. ID NO:26)
15	TSELSKLDATIFAAMKMQWWNPG	(SEQ. ID NO:27)
	VMETGLDLRPILTGMKMDWIPK	(SEQ. ID NO:28)

The sequences of the biotinylated clones from the first library, shown in Table 3, are aligned at the presumably modified K residue. Several clones were present more than once in the set of 20 sequences obtained, so only 15 independent sequences are shown. At some positions in the sequences, no clear consensus is apparent. At other residues, however, clear trends emerge. For example, position -4 (relative to the K residue) was designed to encode V to match that residue of *E. coli* biotin carboxyl carrier protein (GTT codon with each base synthesized 91% as designated, 3% each of the other bases). In spite of this very light mutagenesis, every sequence had a mutation that changed the encoded amino acid to either L or F. L is the residue found at this position in most of the naturally biotinylated sequences from organisms other than *E. coli*, but F was not present in the sequences examined. Residue -3, encoded in the library by a random (NNK) codon, was negatively charged (E or D) in 9 of the 15 sequences. Again, this consensus sequence is similar to that found in the naturally occurring sequences (E or S). The +3 position, however, defines a new consensus not found in the natural sequences. 15 of the 15 peptides had a

hydrophobic residue (W, Y, V, F, or L) at +3, instead of the most commonly found T from the enzyme sequences.

Perhaps the most revealing sequence from the first library was SMWETLNAQKTVLL (SEQ. ID NO:24), which arose from a single base deletion during synthesis or cloning of the library oligonucleotide. This sequence matches only three residues in the enzyme consensus sequence, but does fit the pattern of the other library clones at positions +2 and +3. These results show that the evolutionary constraints on the enzyme sequence result from a combination of factors, only one of which is the ability to be biotinylated.

To define more clearly the consensus sequence for biotinylation, three additional libraries were screened (see Tables 4, 5, and 6, below). Two were based on the pattern from the clones isolated from the first library, and the other consisted simply of a K residue flanked on both sides by 10 random residues. After four rounds of panning, a restriction fragment containing the random region was subcloned from the pool of enriched clones into an MBP (maltose binding protein) expression vector (see U.S. patent application Serial No. 876,288, filed April 29, 1992, incorporated herein by reference). These populations of plasmids were then screened using a colony lift technique involving detection with a streptavidin-alkaline phosphatase conjugate. The biotinylation of several of these clones was confirmed by labeling with ³H-biotin.

The second library was constructed with a random peptide coding sequence defined by xxxIFEAMKMxxxxx (SEQ. ID NO:29); where X is an NNK codon, underlined single residues are codons for the amino acid shown but with a 70/10/10/10 mutagenesis mixture (70% of the base that encodes the amino acid at a particular position in the codon and 10% each of the other three bases), and the codon for K is fixed. The biotinylated sequences isolated and sequenced from this library are shown in Table 4.

Table 4

	LHHILDAQKMVWNHR	(SEQ. ID NO:30)
	PQGIFEAQKMLWRS	(SEQ. ID NO:31)
	LAGTFEALKMAWHEH	(SEQ. ID NO:32)
5	LNAIFEAMKMEYSG	(SEQ. ID NO:33)
	LGGIFEAMKMELRD	(SEQ. ID NO:34)
	LLRTFEAMKMDWRNG	(SEQ. ID NO:35)
	LSTIMEGKMYIQRS	(SEQ. ID NO:36)
	LSDIFEAMKMVYRPC	(SEQ. ID NO:37)
10	LESMLEAMKMOWNPQ	(SEQ. ID NO:38)
	LSDIFDAMKMVYRPQ	(SEQ. ID NO:39)
	LAPFFESMKMVWREH	(SEQ. ID NO:40)
	LKGIFEAMKMEYTAM	(SEQ. ID NO:41)
	LEGIFEAMKMEYSNS	(SEQ. ID NO:42)
15	LLQTFDAMKMEWLPK	(SEQ. ID NO:43)
	VFDILEAQKVVTLRF	(SEQ. ID NO:44)
	LVSMDGDMKMEWCTL	(SEQ. ID NO:45)
	LEPIFEAMKMDWRLE	(SEQ. ID NO:46)
	LKEIFEGMKMEFVKP	(SEQ. ID NO:47)
20	LGGILEAQKMLYRGN	(SEQ. ID NO:48)

The third library was constructed with a random peptide coding sequence defined by xxxxxxIFEAMKMxxxxx (SEQ. ID NO:49); where X is an NNK codon, underlined single residues are codons for the amino acid shown but with a 70/10/10/10 mutagenesis mixture (70% of the base that encodes the amino acid at a particular position in the codon and 10% each of the other three bases), and the codon for K is fixed. The biotinylated sequences isolated and sequenced from this library are shown in Table 5.

Table 5

	RPVLENIFEAMKMEVWKP	(SEQ. ID NO:50)
	RSPIAEIFEAMKMEYRET	(SEQ. ID NO:51)
	QDSIMPIFEAMKMSWHVN	(SEQ. ID NO:52)
5	DGVLFPFIFESMKMIRLET	(SEQ. ID NO:53)
	VSRTMTNFEAMKMIYHDL	(SEQ. ID NO:54)
	DVLLPTVFEAMKMYITK	(SEQ. ID NO:55)
	PNDLERIFDAMKIVTVHS	(SEQ. ID NO:56)
	TRALLEIFDAQKMLYQHL	(SEQ. ID NO:57)
10	RDVHVGIFEAMKMYTVET	(SEQ. ID NO:58)
	GDKLTEIFEAMKIQWTSG	(SEQ. ID NO:59)
	LEGLRAVFESMKMELADE	(SEQ. ID NO:60)
	VADSHDTFAAMKMVWLDLT	(SEQ. ID NO:61)
	GLPLQDILESMTKIVMTSG	(SEQ. ID NO:62)
15	RVPLEAIFEGAKMIWVPNN	(SEQ. ID NO:63)
	PMISHKNFEAMKMLFVPE	(SEQ. ID NO:64)
	KLGLPAMFEAMKMEWHPS	(SEQ. ID NO:65)
	QPSLLSIFEAMKMQASLM	(SEQ. ID NO:66)
	LLELRSNFEAMKMEWQIS	(SEQ. ID NO:67)
20	DEELNQIFEAMKMYPLVHVTK	(SEQ. ID NO:68)

The fourth library was constructed with a random peptide coding sequence defined by xxxxxxxxxxxKxxxxxxxxxxx (SEQ. ID NO:69); where X is an NNK codon, and the codon for K is fixed. The biotinylated sequences obtained and sequenced from this library are shown in Table 6.

Table 6

	SNLVSLLLHSQKILWTDPPQSFG	(SEQ. ID NO:70)
30	LFLHDFLNAQKVELYPVTSSG	(SEQ. ID NO:71)
	SDINALLSTQKIYWAH	(SEQ. ID NO:72)

The biotinylation peptides from these libraries serve to define further the novel consensus sequence for the biotinylation peptides of the present invention. Several features are worth noting. A strong preference for L at position -8 is clear, especially in the second library, which had a shorter random sequence region to the left of the

modified K than any of the other libraries. The other sets of sequences share this preference at -8, but to a lesser extent than in the second library. The L at this position may be more important when there are fewer amino acids connecting the biotinylation domain to the carrier protein. There is no consensus in the naturally occurring sequences at this site.

At other positions, many residues are found and only a general trend is apparent. For example, many residues are found at position -6, but not large hydrophobic residues (L, V, I, W, F, or Y), a tendency that differs from that of the naturally occurring enzymes (L is most frequent). Position +4 contains a wide variety of residues, but with a clear preference for basic amino acids (18 of 56 are R, H, or K) over acidic residues (no D or E).

At position -2, a preference for small size is clear, as only A, G, S, or T are found. Position -1 was biased to be M in all libraries except the fourth library. In these biased libraries, M is found most often, but Q is frequently present. Notably, the mutation from an M codon (ATG) to a Q codon (CAA/G) requires two base changes. In the clones that were unbiased at this position, 4 of 4 clones have Q, indicating that Q might in fact be the preferred residue. The hydrophobic residues M, I, and V are found in almost all of the sequences at position +1. Position +2 is often the natural consensus E but also tends to contain the hydrophobic residues L, V, Y, and I.

To explore the general utility of the biotinylation sequences and to expand their possible uses, a library was made so that the biotinylation peptides would be expressed in a fusion protein at the N-terminus of cytoplasmic MBP. This library was heavily biased in favor of sequences that fit the consensus sequence of the invention, with a random peptide defined by MAXXLXXI(F/L)(E/D)AOK(M/I)EW(H/R)XXXGGS (SEQ. ID NO:73), in which the underlined residues are fixed; the underlined residues are 97/1/1/1 mutagenized codons for the residues shown; and X is an NNK codon. The sequences of positive clones from this library identified by colony lifts are shown in Table 7.

Table 7

5 M A S S L R Q * I L D S Q K M E W R S N A G G S... (SEQ. ID NO:74)
 M A H S L V P I F D A Q K I E W R D P F G G S... (SEQ. ID NO:75)
 M G P D L V N I F E A Q K I E W H P L T G G S... (SEQ. ID NO:76)
 M A F S L R S I L E A Q K M E L R N T P G G S... (SEQ. ID NO:77)
 M A G G L N D I F E A Q K I E W H E D T G G S... (SEQ. ID NO:78)
 M S S Y L A P I F E A Q K I E W H S A Y G G S... (SEQ. ID NO:79)
 M A K A L Q * K I L E A Q K M E W R S H P G G S... (SEQ. ID NO:80)
 10 M A F Q L C K I F Y A Q K M E W H G V G G G S... (SEQ. ID NO:81)
 M A G S L S T I F D A Q K I E W H V G K G G S... (SEQ. ID NO:82)
 M A Q Q L P D I F D A Q K I E W R I A G G G S... (SEQ. ID NO:83)
 M A Q R L F H I L D A Q K I E W H G P K G G S... (SEQ. ID NO:84)
 M A G C L G P I F E A Q K M E W R H F V G G S... (SEQ. ID NO:85)
 M A W S L K P I F D A Q K I E W H S P G G G S... (SEQ. ID NO:86)
 15 M A L G L T R I L D A Q K I E W H R D S G G S... (SEQ. ID NO:87)
 M A G S L R Q I L D A Q K I E W R R P L G G S... (SEQ. ID NO:88)
 M A D R L A Y I L E A Q K M E W H P H K G G S... (SEQ. ID NO:89)

Q* = *supE* suppressed amber codon

20 The biotinylation of several of these clones was confirmed by
 labeling with ³H-biotin. The ability to express functional
 biotinylation sequences free at either end of a protein
 indicates that there is no requirement that either end of the
 25 peptide be free in order to interact with the biotin
 holoenzyme synthetase.

As discussed above, the short, biotinylation
 peptides of the invention can be biotinylated in vivo or in
vitro and can be used for a wide variety of purposes,
 including purification, immobilization, labeling, and
 30 detection of proteins. A few illustrative examples include:
 (1) labeling receptors with biotin at a defined site, so that
 the labeled receptor could be, for instance, bound to
 streptavidin to produce a tetravalent receptor to increase the
 sensitivity of binding assays, such as those described in U.S.
 35 Patent No. 5,143,854, and U.S. patent application Serial No.
 946,239, filed September 16, 1992, each of which is
 incorporated herein by reference; (2) labeling fusion
 proteins containing peptide leads from any screening program,
 so that the labeled fusion proteins can be used to test
 40 binding of the peptide to receptors in a monovalent format (by
 probing with labeled streptavidin after binding occurs) or in

a multivalent format (by prebinding the fusions to labeled streptavidin and then testing binding to receptors or so that the peptides can be immobilized on streptavidin-coated beads or in microtiter wells for probing with receptors, such as protease enzymes, in solution; (3) labeling peptides or proteins directly by growing cells in the presence of tritiated biotin -- with a biotin auxotroph, the peptides could be labeled at a known specific activity to permit quantitative measurements of binding activity; and (4) developing technology for doing enzymatic reactions on surfaces by exposing libraries of variant immobilized sequences to BirA, biotin, and ATP, so that those peptides that were substrates would be biotinylated and could be detected with labeled streptavidin.

This invention also embraces kits which are useful for producing proteins containing biotinylation peptides. Such kits comprise, for instance, a recombinant expression polynucleotide which can be used to produce the peptides of the invention fused to a coding sequence of choice, and directions for using the polynucleotides. DNA expression polynucleotides may be designed to replicate episomally or to integrate into the chromosome of the host cell chosen for expression. Frequently, the DNA polynucleotides of the kit contain a multiple cloning site linked to sequence coding for the peptides of the invention, such that any coding sequence may be inserted in the correct translational reading frame for expression. These kits may be used to produce the peptides of the invention fused to the amino terminus, the carboxyl terminus, or internal to the coding sequence of choice. Within these fusion proteins, the peptides of the invention may be separated from the coding sequences by additional spacer sequences.

Expression of coding sequence will preferably be under control of an inducible promoter; some examples are the lac or tac promoter in E. coli, the gal4 promoter in S. cerevisiae, the glaA promoter in Aspergillus niger, or the murine metallothionein promoter in many mammalian cells. Alternatively, constitutive promoters may be desirable for

certain applications, such as the SV40 early promoter in mammalian cells. For some applications, such as in vitro translation in rabbit reticulocytes, the ability to synthesize RNA in vitro using a RNA polymerase such as that from the bacteriophage SP6 will be needed. In that case, signals for initiation of transcription by both SP6 RNA polymerase and an alternative RNA polymerase can be operably linked to the same expression sequence.

Besides a promoter for initiation of the expression sequences, the polynucleotides of the kits will also preferably contain sequences for transcriptional termination, such as the T7 terminator in E. coli or the SV40 terminator in mammalian cells. Additionally, when the proteins are expressed in mammalian cells, a signal for polyadenylation is desirable, such as the SV40 poly adenylation sequence.

Of course, additional sequences may also be included in the polynucleotides of these kits which will confer additional properties on the proteins produced. For example, a signal sequence which causes the expressed proteins to be secreted from the cell may be incorporated into the polynucleotides. Sequences which serve to link expressed proteins to the membrane, such as a sequence encoding a hydrophobic membrane spanning domain, or an encoded sequence which signals attachment of a glycosyl-phosphatidylinositol membrane anchor to the protein, may be included as part of the expression polynucleotide. The polynucleotides may also encode a sequence recognized by a protease, such as factor Xa, adjacent to the sequence encoding the biotinylation peptides of the invention. One of skill in the art will recognize that these and many other combinations of additional sequences may be advantageous.

Other constituents of the kits may comprise host cells suitable for obtaining expression from the polynucleotide, avidin or streptavidin coupled to a solid support, avidin or streptavidin coupled to a detectable label such as the enzyme horseradish peroxidase, a biotinylation enzyme such as purified BirA, and instructions for analysis and purification of the proteins expressed using these kits.

Preferably, the host cells will express a biotinylating enzyme. Optionally, polynucleotides which, when transformed into host cells, cause the overproduction of biotinylating enzymes may be supplied in the kits, or the host cells provided with the kits may be already modified to over-produce biotinylating enzymes. However, for some applications the absence of biotinylating enzyme in the host cell may be advantageous. For example, the kit user may prefer to biotinylate the expressed fusion proteins in vitro.

Those of skill in the art recognize from the description above that the present invention provides many advantages and more applications than prior art methods for biotinylating proteins. The biotinylation peptides of the invention are small but specific, allowing one to label a protein at a defined site, at either end of or internally to the protein to be labelled. The invention provides an improved immobilization method, allowing one to avoid the use of antibodies and the problems attendant thereto. The high binding affinity of the avidin-biotin interaction provides advantages for labelling, localization, detection, immobilization, and purification methods as well. For instance, one could use the biotinylation peptides of the invention to purify BirA protein or other biotinylation enzymes. The peptides of the invention can serve as the substrate in an assay to screen for the presence of novel biotinylation enzymes. The biotinylation reaction can occur in vivo (where few other proteins are naturally biotinylated) or in vitro, with readily available materials. As can be appreciated from the disclosure above, the present invention has a wide variety of applications. Accordingly, the following examples are offered by way of illustration, not by way of limitation.

Example 1

Library Construction

The peptides on plasmids libraries were made in vector pJS142, a derivative of plasmid pMC5 described in U. S. patent application Serial No. 963,321, filed October 15, 1992,

(GAGGTGGTNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKAAANNKNNKNNKNNKNNKNN
KNNKNNKNNKTAAC TAAGTAAAGC) (SEQ. ID NO:96), respectively, in
place of ON-1396. The fourth library was made with 30 µg of
vector pJS141, which differs from pJS142 only in that the
coding sequence of lacI was altered to encode S, A, and S,
respectively, in place of the C codons normally found at
positions 107, 140, and 281. The library was amplified by
transformation of strain ARI 246 (*lon-11 sulA1 hsdR17*
Δ(ompT-fepC) *ΔclpA319::kan lacI42::Tn10 lacZU118*).

The fifth library was constructed in the vector pBAD/MBP-N, a derivative of pBAD18, see U.S. patent application Serial No. 965,677, filed October 22, 1992, incorporated herein by reference, that places a polylinker and the coding sequence for amino acids 27-393 of MBP downstream from the arabinose-inducible araB promoter. The library was made by ligating annealed ON-1699

(CTAGCTAACTAATGGAGGATACATAAATGgctNNKNNKctgNNKNVKattttNGaNgtc
rAAAatNgaatggcrynNNKNNKNNKGGTGGTAGCC) (SEQ. ID NO:97), where
lower case letters designate 97/1/1/1 mutagenesis; V=A, C, or
G; r=g or a; y=c, t), ON-1700 (TCCTCCATTAGTTAG) (SEQ. ID
NO:98), and ON-1701 (TCGAGGCTACCACC) (SEQ. ID NO:99) to
NheI-XhoI digested pBAD/MBP-N, as described above. The
library was used to transform XL1-Blue (F' proAB lacIq
lacZΔM15 Tn10(tetR) // recA1 endA1 gyrA96 thi hsdR17
supE44 relA1 lac, Stratagene) and screened by colony lifts
as described below.

Panning

About 2 mL of thawed cells in HEK were added to 6 mL of 25 mM HEPES pH 7.5, 0.07 mM EDTA, 8.3% glycerol, 1.25 mg/mL BSA, 0.83 mM DTT, and 0.2 mM PMSF. The cells were lysed for 2 to 4 min. on ice by the addition of 0.15 mL of 10 mg/mL lysozyme (Boehringer Mannheim), and then, 2 mL of 20% lactose

and 0.25 mL of 2 M KCl were added. The supernatant from a 15 min., 27,000 x g centrifugation was added to 0.1 mL of streptavidin-agarose beads (Pierce) in 1 mL HEK, 0.2 M lactose (HEKL), 4.5% BSA, 0.9 mg/mL herring DNA and mixed gently at 4° for 1 hour. The beads were centrifuged and washed 4 times with HEKL buffer, 1% BSA, and 0.1 mg/mL herring DNA at 4° (in later rounds, these washes sometimes contained 10 µM biotin) and then incubated for 30 min. at 4° in the same buffer plus 10 µM biotin. The beads were washed 5 times with HEKL buffer, 1% BSA, twice with HEKL buffer, and once or twice with HEK buffer at 4°. The bound plasmids were eluted with 35 mM HEPES pH 7.5, 0.1 mM EDTA, 200 mM KCl, 1 mM IPTG, 10 µg/mL self-annealed ON-413 (GAATTCAATTGTGAGCGCTCACAATTGAATTC) (SEQ. ID NO:100) for 30 min. at room temperature, precipitated with isopropanol, and then used to electrotransform either ARI 280 or ARI 298 (*lon-11 sulA1 hsdR17 Δ(ompT-fepC) ΔclpA319::kan ΔlacI lacZU118 recA::cat cytR*) for amplification.

Example 3

Subcloning into MBP Vector

Plasmids recovered from panning were digested with BspEI and ScaI, and a fragment containing the peptide coding sequence was subcloned into AgeI, ScaI, digested plasmid pELM3, a derivative of pMALc2, which is available from New England Biolabs, designed to accept inserts of coding sequence from pJS142. The transferred fragment encodes GGG-peptide and is linked to the MBP coding region through sequence encoding N₁₀LGIEGRT. The MBP is retained in the cytoplasm due to the lack of a signal sequence.

Example 4

Labeling with ³H-biotin

Cells were grown at 37° overnight in minimal medium E (Davis, 1980, Advanced Bacterial Genetics (CSH Press)) with 0.4% glycerol, 0.1% vitamin assay casamino acids (Difco), 1 µg/mL thiamine, and 50 µg/mL ampicillin. The cultures were diluted 1/10 in the same medium, grown for several hours, and then added to an equal volume of medium containing 2 µCi/mL.

³H-Biotin (Amersham) and 0.6 mM IPTG (for pELM3 clones) or 0.4% L-arabinose instead of glycerol (for pBAD/MBP-N clones). Growth was continued for an additional several hours, and then the cells were harvested and lysed with SDS protein gel buffer. The samples were run on a 4-20% gradient acrylamide gel, and fluorographed using Amplify (Amersham) and X-ray film.

Example 5

Colony Lifts

Colony lifts were performed in duplicates essentially as described (Sambrook, 1989, Molecular Cloning: A Laboratory Manual (CSH Press)), except that the inducing plates contained 10 μ M biotin and 0.3 mM IPTG (for pELM3 clones) or 0.2% L-arabinose (for pBAD/MBP-N clones). The blocking agent was 5% BSA, and the probe was 1/5000 diluted streptavidin-alkaline phosphatase conjugate (Gibco BRL).

Example 6

Overexpression of BirA

The birA gene was cloned under the control of inducible promoters on two different plasmids. The birA gene was amplified from the plasmid pBA22 (see Barker and Campbell, 1981, J. Mol. Biol. 146:469-492) using primers ON-1589 (SEQ. ID NO:101) (5'TAC AGT GCT AGC TAA CTA ATG GAG GAT ACA TAA ATG AAG GAT AAC ACC GTG CCA CTG 3') and ON-1590 (SEQ. ID NO:102) (5' GTA TCA GAG CTC TTA TTT TTC TGC ACT ACG CAG GGA 3') in a polymerase chain reaction (PCR). The fragment was digested with SacI and NheI and cloned into SacI, NheI digested plasmid pJS100, placing birA under control of the araBAD promoter. The resulting plasmid, called pJS170, contains a pBR322-derived replication origin, an ampicillin resistance gene, and the araC gene, which encodes a regulator of the araBAD promoter. Induction of birA expression from this plasmid in LB + 0.2% arabinose allows expression of large amounts of BirA protein.

The birA gene fragment was also subcloned into SacI, SpeI digested plasmid pIQCAT-LC9. This places birA under

control of the tac promoter, which is inducible with IPTG. This plasmid, called pJS169, also contains a p15A replication origin, a chloramphenicol resistance gene, and the lacIQ allele of the lacI gene, which encodes a repressor of the lac or tac promoters. The p15A replication origin permits this plasmid to replicate in the same cell as pBR322-derived plasmids. Thus, BirA can be overexpressed in the same cell that is expressing the biotinylation target. Cells carrying pJS169 grown in LB + 0.3 mM IPTG overexpress BirA to a lesser extent than cells carrying pJS170 induced with 0.2% arabinose.

Example 7

Enhanced Biotinylation in an *E. coli* strain over-producing BirA

The efficiency of MBP-peptide fusion biotinylation was determined under two growth condition using a band shift assay. This assay was performed by mixing deglycosylated avidin (UltraAvidin, Leinco Technologies) with a crude cell lysate from an *E. coli* strain that overexpressed the MBP-peptide fusion. The mix was electrophoresed on a 4-20% acrylamide nondenaturing gel compared to the lysate without avidin. Comparison of the two lanes permitted quantitation of the efficiency of biotinylation by observation of the band shift caused by the added avidin. Fusion proteins expressed in a strain carrying pJS169 (with birA induced with 0.3 mM IPTG) in LB media containing 10 μ M biotin were biotinylated to a greater extent than those expressed in the absence of extra BirA and added biotin.

Example 8

Biotinylation of recombinant proteins in vitro

1. BirA overexpression and purification

BirA can be purified either by published procedures (see Buoncristiani and Otsuka, 1988, J. Biol. Chem. 263(2):1013-1016), or by the following procedure.

A single colony of *E. coli* strain BL21 transformed with pJS169 was grown overnight in 50 mls of LB + ampicillin. This culture was innoculated 1:100 into 1 liter of LB + ampicillin and grown at 37°C with shaking until the OD₆₀₀ =

0.5. After induction with 0.4 mM IPTG, the cells were grown an additional 4 hours, harvested by centrifugation, resuspended in 20mM Tris-HCl pH 7.4 + 5mM DTT (TD5), and lysed by sonication. Cellular debris was removed by centrifugation and the supernatant diluted to 100 ml total volume with TD5 buffer.

Crude supernatant was loaded onto a 10 ml Blue Sepharose FF column (Pharmacia) and washed through with TD (20mM Tris-HCl pH 7.4) until the A280 of the column flow-through was about 0. This column was eluted with a 100 ml gradient of 0-1.5M NaCl in TD and 2 ml fractions collected. BirA-containing fractions were pooled and dialyzed against TD until the NaCl concentration was about 15 mM.

The dialysate was concentrated using an Amicon YM30, and then loaded over 5ml S Sepharose FF column (Pharmacia) and washed through with TD1 (20mM Tris-HCl pH 7.4 + 1mM DTT). Protein was eluted with a 50 ml gradient of 0-350 mM NaCl in TD1.

BirA-containing fractions were pooled, bound to a Biotin-sepharose column, washed with TD1/150mM NaCl, and eluted with TD1/ 150mM NaCl +2mM biotin. BirA-containing fractions were dialyzed over YM30 against TD1/150mM NaCl to a final volume of 10 ml.

2. Biotinylation in vitro using purified BirA enzyme.

Proteins fused to one of the peptides of the invention were biotinylated in vitro at 37°C in a buffer containing: RPMI medium 1640 (Gibco-BRL) supplemented with 5mM ATP, 5mM MgCl₂, and 10μM biotin.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity and understanding, it will be apparent that certain changes and modifications may be practiced within the scope of the appended claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Affymax Technologies N.V.
- (ii) TITLE OF INVENTION: Biotinylation of Proteins
- (iii) NUMBER OF SEQUENCES: 102
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Townsend and Townsend Khourie and Crew
 - (B) STREET: One Market Plaza, Steuart Tower
 - (C) CITY: San Francisco
 - (D) STATE: California
 - (E) COUNTRY: USA
 - (F) ZIP: 94105
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: PCT
 - (B) FILING DATE:
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08/099,991
 - (B) FILING DATE: 30-JUL-1993
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Smith, William M.
 - (B) REGISTRATION NUMBER: 30,223
 - (C) REFERENCE/DOCKET NUMBER: 1038.1
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 415-326-2400
 - (B) TELEFAX: 415-326-2422

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Region
- (B) LOCATION: one-of(2, 3)
- (D) OTHER INFORMATION: /note= "Xaa at position 2 is any amino acid; at position 3 is any amino acid other than Leu, Val, Ile, Trp, Phe or Tyr."

(ix) FEATURE:

- (A) NAME/KEY: Region
- (B) LOCATION: one-of(5, 6, 7, 8)
- (D) OTHER INFORMATION: /note= "Xaa at position 5 is Phe or Leu; at position 6 is Glu or Asp; at position 7 is Ala, Gly, Ser or Thr; at position 8 is Gln or

(ix) FEATURE:

- (A) NAME/KEY: Region
- (B) LOCATION: one-of(10, 11, 12, 13)
- (D) OTHER INFORMATION: /note= "Xaa at position 10 is Ile, Met or Val; at position 11 is Glu, Leu, Val, Tyr or Ile; at position 12 is Trp, Tyr Val, Phe, Leu

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Leu Xaa Xaa Ile Xaa Xaa Xaa Xaa Lys Xaa Xaa Xaa Xaa
1 5 10

(2) INFORMATION FOR SEQ ID NO:2:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Xaa Ile Val Xaa Ala Met Lys Met Xaa
1 5

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Gly Gln Thr Val Leu Val Leu Glu Ala Met Lys Met Glu Thr Glu Ile
1 5 10 15
Asn Ala Pro Thr Asp Gly
 20

(A) LENGTH: 22 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Gly Glu Val Leu Leu Ile Leu Glu Ala Met Lys Met Glu Thr Glu Ile
1 5 10 15
Arg Ala Ala Gln Ala Gln
20

(A) LENGTH: 22 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Gly Gln Cys Phe Ala Glu Ile Glu Val Met Lys Met Val Met Thr Leu
1 5 10 15

Thr Ala Gly Glu Ser Gly
20

CHARACTERISTICS:

- (A) LENGTH: 22 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Gly Asn Thr Leu Cys Ile Val Glu Ala Met Lys Met Met Asn Gln Ile
1 5 10 15
Glu Ala Asp Lys Ser Gly
20

(A) LENGTH: 22 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Gly Gln Pro Val Ala Val Leu Ser Ala Met Lys Met Glu Met Ile Ile
1 5 10 15
Ser Ser Pro Ser Asp Gly
20

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 22 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Gly Gln Pro Leu Cys Val Leu Ser Ala Met Lys Met Glu Thr Val Val
1 5 10 15
Thr Ser Pro Met Glu Gly
 20

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Gly Gln Pro Leu Val Leu Ser Ala Met Lys Met Glu Thr Val Val Thr
1 5 10 15
Ser Pro Val Thr Glu
20

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Gly Ala Pro Leu Val Leu Ser Ala Met Lys Met Glu Thr Val Val Thr
1 5 10 15

Ala Pro Arg

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Gly Gln Glu Ile Cys Val Ile Glu Ala Met Lys Met Gln Asn Ser Met
1 5 10 15
Thr Ala Gly Lys Thr Gly
 20

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Gly	Gln	Pro	Val	Leu	Val	Leu	Glu	Ala	Met	Lys	Met	Glu	His	Val	Val	
1				5					10					15		
Lys	Ala	Pro	Ala	Asn	Gly											
					20											

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Ala Met Lys Met

1

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 26 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Leu	Glu	Glu	Val	Asp	Ser	Thr	Ser	Ser	Ala	Ile	Phe	Asp	Ala	Met	Lys
1				5					10					15	
Met	Val	Trp	Ile	Ser	Pro	Thr	Glu	Phe	Arg						
			20						25						

(A) LENGTH: 27 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Gln Gly Asp Arg Asp Glu Thr Leu Pro Met Ile Leu Arg Ala Met Lys
1 5 10 15

Met Glu Val Tyr Asn Pro Gly Gly His Glu Lys
20 25

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Ser	Lys	Cys	Ser	Tyr	Ser	His	Asp	Leu	Lys	Ile	Phe	Glu	Ala	Gln	Lys
1				5					10					15	
Met	Leu	Val	His	Ser	Tyr	Leu	Arg	Val	Met	Tyr	Asn	Tyr			
			20					25							

(A) LENGTH: 22 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Met Ala Ser Ser Asp Asp Gly Leu Leu Thr Ile Phe Asp Ala Thr Lys
1 5 10 15

Met Met Phe Ile Arg Thr
20

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Ser	Tyr	Met	Asp	Arg	Thr	Asp	Val	Pro	Thr	Ile	Leu	Glu	Ala	Met	Lys
1				5				10						15	
Met	Glu	Leu	His	Thr	Thr	Pro	Trp	Ala	Cys	Arg					
			20					25							

(2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 27 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Ser Val Val Pro Glu Pro Gly Trp Asp Gly Pro Phe Glu Ser Met Lys
1 5 10 15

Met Val Tyr His Ser Gly Ala Gln Ser Gly Gln
 20 25

(A) LENGTH: 21 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Ser Phe Pro Pro Ser Leu Pro Asp Lys Asn Ile Phe Glu Ala Met Lys
1 5 10 15

Met Tyr Val Ile Thr
20

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Val	Arg	His	Leu	Pro	Pro	Pro	Leu	Pro	Ala	Leu	Phe	Asp	Ala	Met	Lys
1				5					10					15	
Met	Glu	Phe	Val	Thr	Ser	Val	Gln	Phe							
			20					25							

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Asp Met Thr Met Pro Thr Gly Met Thr Lys Ile Phe Glu Ala Met Lys
1 5 10 15
Met Glu Val Ser Thr
20

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

Ala	Thr	Ala	Gly	Pro	Leu	His	Glu	Pro	Asp	Ile	Phe	Leu	Ala	Met	Lys
1				5				10						15	
Met	Glu	Val	Val	Asp	Val	Thr	Asn	Lys	Ala	Gly	Gln				
		20						25							

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Ser Met Trp Glu Thr Leu Asn Ala Gln Lys Thr Val Leu Leu
1 5 10

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

Ser His Pro Ser Gln Leu Met Thr Asn Asp Ile Phe Glu Gly Met Lys
1 5 10 15

Met Leu Tyr His
20

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

Ser	Ile	Glu	Arg	Gly	Gly	Ser	Thr	His	Lys	Ile	Leu	Ala	Ala	Met	Lys
1				5					10					15	
Met	Tyr	Gln	Val	Ser	Thr	Pro	Ser	Cys	Ser						
			20					25							

(A) LENGTH: 23 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

Thr Ser Glu Leu Ser Lys Leu Asp Ala Thr Ile Phe Ala Ala Met Lys
1 5 10 15

Met Gln Trp Trp Asn Pro Gly
20

(2) INFORMATION FOR SEQ ID NO:28:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

Val Met Glu Thr Gly Leu Asp Leu Arg Pro Ile Leu Thr Gly Met Lys
1 5 10 15
Met Asp Trp Ile Pro Lys
20

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Region
- (B) LOCATION: one-of(1, 2, 3, 11, 12, 13, 14, 15)
- (D) OTHER INFORMATION: /note= "Xaa is any amino acid coded for by the NNK codon."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

Xaa	Xaa	Xaa	Ile	Phe	Glu	Ala	Met	Lys	Met	Xaa	Xaa	Xaa	Xaa	Xaa
1				5					10					15

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

Leu His His Ile Leu Asp Ala Gln Lys Met Val Trp Asn His Arg
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

Pro Gln Gly Ile Phe Glu Ala Gln Lys Met Leu Trp Arg Ser
1 5 10

(2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

Leu Ala Gly Thr Phe Glu Ala Leu Lys Met Ala Trp His Glu His
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

Leu Asn Ala Ile Phe Glu Ala Met Lys Met Glu Tyr Ser Gly
1 5 10

(2) INFORMATION FOR SEQ ID NO:34:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 14 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

Leu Gly Gly Ile Phe Glu Ala Met Lys Met Glu Leu Arg Asp
1 5 10

(2) INFORMATION FOR SEQ ID NO:35:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

Leu	Leu	Arg	Thr	Phe	Glu	Ala	Met	Lys	Met	Asp	Trp	Arg	Asn	Gly
1				5				10					15	

(2) INFORMATION FOR SEQ ID NO:36:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 15 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

Leu	Ser	Thr	Ile	Met	Glu	Gly	Met	Lys	Met	Tyr	Ile	Gln	Arg	Ser
1				5					10					15

(2) INFORMATION FOR SEQ ID NO:37:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

Leu	Ser	Asp	Ile	Phe	Glu	Ala	Met	Lys	Met	Val	Tyr	Arg	Pro	Cys
1				5					10					15

(2) INFORMATION FOR SEQ ID NO:38:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 15 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

Leu	Glu	Ser	Met	Leu	Glu	Ala	Met	Lys	Met	Gln	Trp	Asn	Pro	Gln
1				5					10					15

(2) INFORMATION FOR SEQ ID NO:39:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

Leu	Ser	Asp	Ile	Phe	Asp	Ala	Met	Lys	Met	Val	Tyr	Arg	Pro	Gln
1				5				10						15

(2) INFORMATION FOR SEQ ID NO:40:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

Leu	Ala	Pro	Phe	Phe	Glu	Ser	Met	Lys	Met	Val	Trp	Arg	Glu	His
1				5					10					15

(2) INFORMATION FOR SEQ ID NO:41:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 15 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

Leu Lys Gly Ile Phe Glu Ala Met Lys Met Glu Tyr Thr Ala Met
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:42:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 15 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

Leu	Glu	Gly	Ile	Phe	Glu	Ala	Met	Lys	Met	Glu	Tyr	Ser	Asn	Ser
1				5					10					15

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

Leu Leu Gln Thr Phe Asp Ala Met Lys Met Glu Trp Leu Pro Lys
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:44:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

Val Phe Asp Ile Leu Glu Ala Gln Lys Val Val Thr Leu Arg Phe
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

Leu Val Ser Met Phe Asp Gly Met Lys Met Glu Trp Lys Thr Leu
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

Leu	Glu	Pro	Ile	Phe	Glu	Ala	Met	Lys	Met	Asp	Trp	Arg	Leu	Glu
1				5					10					15

(2) INFORMATION FOR SEQ ID NO:47:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

Leu Lys Glu Ile Phe Glu Gly Met Lys Met Glu Phe Val Lys Pro
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:48:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

Leu Gly Gly Ile Glu Ala Gln Lys Met Leu Leu Tyr Arg Gly Asn
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

(A) NAME/KEY: Region

(B) LOCATION: one-of(1, 2, 3, 4, 5, 6, 14, 15, 16, 17, 18)

(D) OTHER INFORMATION: /note= "Xaa is any amino acid coded for by the NNK codon."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Ile	Phe	Glu	Ala	Met	Lys	Met	Xaa	Xaa	Xaa
1				5					10					15	
Xaa Xaa															

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

Arg Pro Val Leu Glu Asn Ile Phe Glu Ala Met Lys Met Glu Val Trp
1 5 10 15

Lys Pro

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

Arg	Ser	Pro	Ile	Ala	Glu	Ile	Phe	Glu	Ala	Met	Lys	Met	Glu	Tyr	Arg
1				5				10						15	
Glu Thr															

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

Gln	Asp	Ser	Ile	Met	Pro	Ile	Phe	Glu	Ala	Met	Lys	Met	Ser	Trp	His
1				5					10					15	
Val Asn															

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

Asp	Gly	Val	Leu	Phe	Pro	Ile	Phe	Glu	Ala	Met	Lys	Met	Ile	Arg	Leu
1				5				10						15	
Glu Thr															

(2) INFORMATION FOR SEQ ID NO:54:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 18 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

Val	Ser	Arg	Thr	Met	Thr	Asn	Phe	Glu	Ala	Met	Lys	Met	Ile	Tyr	His
1				5					10					15	
Asp Leu															

(2) INFORMATION FOR SEQ ID NO:55:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:55:

Asp Val Leu Leu Pro Thr Val Phe Glu Ala Met Lys Met Tyr Ile Thr
1 5 10 15

Lys

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

Pro	Asn	Asp	Leu	Glu	Arg	Ile	Phe	Asp	Ala	Met	Lys	Ile	Val	Thr	Val
1				5					10					15	

His Ser

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

Thr	Arg	Ala	Leu	Leu	Glu	Ile	Phe	Asp	Ala	Gln	Lys	Met	Leu	Tyr	Gln
1			5						10					15	
His Leu															

(A) LENGTH: 18 amino acids

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

Arg Asp Val His Val Gly Ile Phe Glu Ala Met Lys Met Tyr Thr Val
1 5 10 15

Glu Thr

(2) INFORMATION FOR SEQ ID NO:59:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

Gly Asp Lys Leu Thr Glu Ile Phe Glu Ala Met Lys Ile Gln Trp Thr
1 5 10 15
Ser Gly

(A) LENGTH: 18 amino acids

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

Leu Glu Gly Leu Arg Ala Val Phe Glu Ser Met Lys Met Glu Leu Ala
1 5 10 15

Asp Glu

(2) INFORMATION FOR SEQ ID NO:61:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

Val Ala Asp Ser His Asp Thr Phe Ala Ala Met Lys Met Val Trp Leu
1 5 10 15
Asp Thr

(2) INFORMATION FOR SEQ ID NO:62:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 18 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

Gly Leu Pro Leu Gln Asp Ile Leu Glu Ser Met Lys Ile Val Met Thr
1 5 10 15
Ser Gly

(A) LENGTH: 19 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

Arg Val Pro Leu Glu Ala Ile Phe Glu Gly Ala Lys Met Ile Trp Val
1 5 10 15
Pro Asn Asn

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

Pro Met Ile Ser His Lys Asn Phe Glu Ala Met lys Met Lys Phe Val
1 5 10 15

Pro Glu

(2) INFORMATION FOR SEQ ID NO:65:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

Lys Leu Gly Leu Pro Ala Met Phe Glu Ala Met Lys Met Glu Trp His
1 5 10 15
Pro Ser

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

Gln Pro Ser Leu Leu Ser Ile Phe Glu Ala Met Lys Met Gln Ala Ser
1 5 10 15
Leu Met

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

Leu Leu Glu Leu Arg Ser Asn Phe Glu Ala Met Lys Met Glu Trp Gln
1 5 10 15

Ile Ser

(2) INFORMATION FOR SEQ ID NO:68:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

Asp Glu Glu Leu Asn Gln Ile Phe Glu Ala Met Lys Met Tyr Pro Leu
1 5 10 15
Val His Val Thr Lys
 20

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

(A) NAME/KEY: Region

(B) LOCATION: one-of(1..10, 12..21)

(D) OTHER INFORMATION: /note= "Xaa is any amino acid coded for by the NNK codon."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Lys Xaa Xaa Xaa Xaa Xaa
1 5 10 15

Xaa Xaa Xaa Xaa Xaa
20

(A) LENGTH: 21 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

Ser Asn Leu Val Ser Leu Leu His Ser Gln Lys Ile Leu Trp Thr Asp
1 5 10 15

Pro Gln Ser Phe Gly
20

(A) LENGTH: 21 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

Leu Phe Leu His Asp Phe Leu Asn Ala Gln Lys Val Glu Leu Tyr Pro
1 5 10 15

Val Thr Ser Ser Gly
20

(2) INFORMATION FOR SEQ ID NO:72:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 16 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

Ser Asp Ile Asn Ala Leu Leu Ser Thr Gln Lys Ile Tyr Trp Ala His
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:73:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: one-of(3, 4, 6, 7, 18, 19, 20)
- (D) OTHER INFORMATION: /note= "Xaa is any amino acid coded for by the NNK codon."

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: one-of(9, 10, 14, 17)
- (D) OTHER INFORMATION: /note= "Xaa at position 9 is Phe or Leu; at position 10 is Glu or Asp; at position 14 is Met or Ile; at position 17 is His or Arg."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

Met	Ala	Xaa	Xaa	Leu	Xaa	Xaa	Ile	Xaa	Xaa	Ala	Gln	Lys	Xaa	Glu	Trp
1				5				10						15	
Xaa Xaa Xaa Xaa Gly Gly Ser															
20															

(2) INFORMATION FOR SEQ ID NO:74:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

Met	Ala	Ser	Ser	Leu	Arg	Gln	Ile	Leu	Asp	Ser	Gln	Lys	Met	Glu	Trp
1				5				10						15	
Arg Ser Asn Ala Gly Gly Ser															
20															

(A) LENGTH: 23 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

Met Ala His Ser Leu Val Pro Ile Phe Asp Ala Gln Lys Ile Glu Trp
1 5 10 15

Arg Asp Pro Phe Gly Gly Ser
20

(2) INFORMATION FOR SEQ ID NO:76:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

Met Gly Pro Asp Leu Val Asn Ile Phe Glu Ala Gln Lys Ile Glu Trp
1 5 10 15
His Pro Leu Thr Gly Gly Ser
20

(A) LENGTH: 23 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

Met Ala Phe Ser Leu Arg Ser Ile Leu Glu Ala Gln Lys Met Glu Leu
1 5 10 15

Arg Asn Thr Pro Gly Gly Ser
20

(2) INFORMATION FOR SEQ ID NO:78:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

Met	Ala	Gly	Gly	Leu	Asn	Asp	Ile	Phe	Glu	Ala	Gln	Lys	Ile	Glu	Trp
1				5					10					15	
His Glu Asp Thr Gly Gly Ser															
20															

(2) INFORMATION FOR SEQ ID NO:79:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

Met Ser Ser Tyr Leu Ala Pro Ile Phe Glu Ala Gln Lys Ile Glu Trp
1 5 10 15
His Ser Ala Tyr Gly Gly Ser
20

(2) INFORMATION FOR SEQ ID NO:80:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

Met	Ala	Lys	Ala	Leu	Gln	Lys	Ile	Leu	Glu	Ala	Gln	Lys	Met	Glu	Trp
1				5				10						15	
Arg Ser His Pro Gly Gly Ser															
20															

(2) INFORMATION FOR SEQ ID NO:81:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

Met	Ala	Phe	Gln	Leu	Cys	Lys	Ile	Phe	Tyr	Ala	Gln	Lys	Met	Glu	Trp
1				5					10					15	
His Gly Val Gly Gly Gly Ser															
20															

(2) INFORMATION FOR SEQ ID NO:82:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

Met	Ala	Gly	Ser	Leu	Ser	Thr	Ile	Phe	Asp	Ala	Gln	Lys	Ile	Glu	Trp
1				5					10					15	
His Val Gly Lys Gly Gly Ser															
20															

(2) INFORMATION FOR SEQ ID NO:83:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

Met	Ala	Gln	Gln	Leu	Pro	Asp	Ile	Phe	Asp	Ala	Gln	Lys	Ile	Glu	Trp
1				5					10					15	
Arg Ile Ala Gly Gly Gly Ser															
20															

(A) LENGTH: 23 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

Met Ala Gln Arg Leu Phe His Ile Leu Asp Ala Gln Lys Ile Glu Trp
1 5 10 15
His Gly Pro Lys Gly Gly Ser
20

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

Met Ala Gly Cys Leu Gly Pro Ile Phe Glu Ala Gln Lys Met Glu Trp
1 5 10 15
Arg His Phe Val Gly Gly Ser
20

(2) INFORMATION FOR SEQ ID NO:86:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

Met	Ala	Trp	Ser	Leu	Lys	Pro	Ile	Phe	Asp	Ala	Gln	Lys	Ile	Glu	Trp
1				5					10					15	
His Ser Pro Gly Gly Gly Ser															
20															

(A) LENGTH: 23 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

Met Ala Leu Gly Leu Thr Arg Ile Leu Asp Ala Gln Lys Ile Glu Trp
1 5 10 15
His Arg Asp Ser Gly Gly Ser
20

(2) INFORMATION FOR SEQ ID NO:88:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 23 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

Met Ala Gly Ser Leu Arg Gln Ile Leu Asp Ala Gln Lys Ile Glu Trp
1 5 10 15
Arg Arg Pro Leu Gly Gly Ser
 20

(A) LENGTH: 23 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

Met Ala Asp Arg Leu Ala Tyr Ile Leu Glu Ala Gln Lys Met Glu Trp
1 5 10 15

His Pro His Lys Gly Gly Ser
20

(2) INFORMATION FOR SEQ ID NO:90:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 15 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

Val	Val	His	Gly	Glu	Gln	Val	Gly	Gly	Glu	Ala	Ser	Gly	Gly	Gly
1				5				10					15	

(2) INFORMATION FOR SEQ ID NO:91:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 103 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (oligonucleotide)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:

GAGGTGGTNN KNNKNNKNNK NNNKNNKNNK NNNKNNKAT CGTTNNKGCT ATGAAAATGN	60
NNNNKNNKNN KNNKNNKNNK NNNKNNKNNKT AACTAAGTAA AGC	103

(2) INFORMATION FOR SEQ ID NO:92:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (oligonucleotide)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:

ACCACCTCCG G

(2) INFORMATION FOR SEQ ID NO:93:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 11 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (oligonucleotide)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:93:

TTACTTAGTT A

11

(2) INFORMATION FOR SEQ ID NO:94:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 67 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (oligonucleotide)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

GAGGTGGTNN KNNKNNKATC TTTGAAGCTA TGAAAATGNN KNNKNNKNNK NNKTAACATA
GTAAAGC

60

67

(2) INFORMATION FOR SEQ ID NO:95:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 76 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (oligonucleotide)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:

GAGGTGGTNN KNNKNNKNNK NNKNNKATCT TTGAAGCTAT GAAAATGNNK NNKNNKNNKN
NKTAATAAG TAAAGC

60

76

(2) INFORMATION FOR SEQ ID NO:96:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 85 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (oligonucleotide)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:

GAGGTGGTNN KNNKNNKNNK NNNKNNKNNK NNNKNNKAA ANNNKNNKNNK NNNKNNKNNK
NNNNKNNKNN KTAACATAAGT AAAGC

60

85

(2) INFORMATION FOR SEQ ID NO:97:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 96 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (oligonucleotide)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:97:

CTAGCTAACT AATGGAGGAT ACATAAATGG CTNNKNNKCT GNNKNVKATT TTNGANGCTC	60
ARAAAATNGA ATGGCRYNNK NNKNNKGGTG GTAGCC	96

(2) INFORMATION FOR SEQ ID NO:98:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 15 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (oligonucleotide)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:98:

TCCTCCATTA GTTAG

(2) INFORMATION FOR SEQ ID NO:99:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (oligonucleotide)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:

TCGAGGCTAC CACC

(2) INFORMATION FOR SEQ ID NO:100:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 32 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA (oligonucleotide)

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:

GAATTCATT GTGAGCGCTC ACAATTGAAT TC

(2) INFORMATION FOR SEQ ID NO:101:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 57 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (primer)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:101:

TACAGTGCTA GCTAACTAAT GGAGGATACA TAAATGAAGG ATAACACCGT GCCACTG

57

(2) INFORMATION FOR SEQ ID NO:102:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 36 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA (primer)

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:102:

GTATCAGAGC TCTTATTTTT CTGCACTACG CAGGGA

WHAT IS CLAIMED IS:

1 1. A method for biotinylating a protein, said
2 method comprising:

3 (a) constructing a recombinant DNA expression vector
4 that encodes a fusion protein comprising said protein and a
5 biotinylation peptide less than 50 amino acids in length;

6 (b) transforming a recombinant host cell with said
7 vector; and

8 (c) culturing said host cell in the presence of
9 biotin or a biotin analogue and under conditions such that
10 said fusion protein and a biotinylation enzyme are expressed,
11 resulting in biotinylation of said fusion protein.

1 2. The method of Claim 1, wherein the host cell is
2 E. coli.

1 3. The method of Claim 1, wherein said
2 biotinylation peptide comprises an amino acid sequence defined
3 by: LX₁X₂IX₃X₄X₅X₆KX₇X₈X₉X₁₀ (SEQ. ID NO:1), wherein X₁ is any
4 amino acid; X₂ is any amino acid other than L, V, I, W, F, Y;
5 X₃ is F or L; X₄ is E or D; X₅ is A, G, S, or T; X₆ is Q or M;
6 X₇ is I, M, or V; X₈ is E, L, V, Y, or I; X₉ is W, Y, V, F, L,
7 or I; and X₁₀ is R, H, or any amino acid other than D or E..

1 4. A method for biotinylating a protein, said
2 method comprising:

3 (a) constructing a recombinant DNA expression vector
4 that encodes a fusion protein comprising said protein and a
5 biotinylation peptide less than 50 amino acids in length;

6 (b) producing said fusion protein encoded by said
7 vector either by transforming a recombinant host cell with
8 said vector and culturing host cells transformed with the
9 vector or by incubating said vector in a cell-free
10 transcription and translation system; and

11 (c) incubating said fusion protein in a reaction
12 mixture comprising biotin or a biotin analogue and a

13 biotinylation enzyme, resulting in biotinylation of said
14 fusion protein.

1 5. A recombinant DNA vector that comprises a
2 nucleic acid that encodes a biotinylation peptide less than 50
3 amino acids in length.

1 6. The recombinant DNA vector of Claim 5, wherein
2 said biotinylation peptide is $LX_1X_2IX_3X_4X_5X_6KX_7X_8X_9X_{10}$ (SEQ. ID
3 NO:1), where X_1 is any amino acid; X_2 is any amino acid other
4 than L, V, I, W, F, Y; X_3 is F or L; X_4 is E or D; X_5 is A, G,
5 S, or T; X_6 is Q or M; X_7 is I, M, or V; X_8 is E, L, V, Y, or
6 I; X_9 is W, Y, V, F, L, or I; and X_{10} is preferably R, H, or
7 any amino acid other than D or E.

1 7. The recombinant DNA vector of Claim 5, wherein
2 said biotinylation peptide is selected from the group
3 consisting of LEEVDSTSSAIFDAMKMVWISPTEFR (SEQ. ID NO:14),
4 QGDRDETLPILRAMKMEVYNPGGHEK (SEQ. ID NO:15),
5 SKCSYSHDLKIFEAQKMLVHSYLRVMYNY (SEQ. ID NO:16),
6 MASSDDGLLTIFDATKMMFIRT (SEQ. ID NO:17),
7 SYMDRTDVPTILEAMKMEVHTTPWACR (SEQ. ID NO:18),
8 SFPPSLPDKNIFEAMKMYVIT (SEQ. ID NO:19),
9 SVVPEPGWDGPFESMKMVYHSGAQSGQ (SEQ. ID NO:20),
10 VRHLPPPLPALFDAMKMEFVTSVQF (SEQ. ID NO:21),
11 DMTMPTGMTKIFEAMKMEVST (SEQ. ID NO:22),
12 ATAGPLHEPDIFLAMKMEVVDVTNKAGQ (SEQ. ID NO:23), SMWETLNAQKTVLL
13 (SEQ. ID NO:24), SHPSQLMTNDIFEGMKMLYH (SEQ. ID NO:25),
14 TSELSKLDATIFAAMKMQWWNPG (SEQ. ID NO:27),
15 VMETGLDLRPILTGMKMDWIPK (SEQ. ID NO:28), LHHILDAQKMVWNHR (SEQ.
16 ID NO:30), PQGIFEAQKMLWRS (SEQ. ID NO:31), LAGTFEALKMAWHEH
17 (SEQ. ID NO:32), LNAIFEAMKMEYSG (SEQ. ID NO:33),
18 LGGIFEAMKMELRD (SEQ. ID NO:34), LLRTFEAMKMDWRNG (SEQ. ID
19 NO:35), LSTIMEGKMKYIORS (SEQ. ID NO:36), LSDIFEAMKMVYRPC (SEQ.
20 ID NO:37), LESMLEAMKMQWNPQ (SEQ. ID NO:38), LSDIFDAMKMVYRPQ
21 (SEQ. ID NO:39), LAPFFESMKMVWREH (SEQ. ID NO:40),
22 LKGIFEAMKMEYTAM (SEQ. ID NO:41), LEGIFEAMKMEYSNS (SEQ. ID
23 NO:42), LLQTFDAMKMEWLPK (SEQ. ID NO:43), VFDILEAQKVVTLR (SEQ.

24 ID NO:44), LVSMFDGMKMEWKTL (SEQ. ID NO:45), LEPIFEAMKMDWRLE
25 (SEQ. ID NO:46), LKEIFEGMKMEFVKP (SEQ. ID NO:47),
26 LGGILEAQKMLYRGN (SEQ. ID NO:48), RPVLENIFEAMKMEVWKP (SEQ. ID
27 NO:50), RSPIAEIFEAMKMEYRET (SEQ. ID NO:51), QDSIMPIFEAMKMSWHVN
28 (SEQ. ID NO:52), DGVLFPIFESMKMIRLET (SEQ. ID NO:53),
29 VSRTMTNFEAMKMIYHDL (SEQ. ID NO:54), DVLLPTVFEAMKMYITK (SEQ. ID
30 NO:55), PNDLERIFDAMKIVTVHS (SEQ. ID NO:56), TRALLEIFDAQKMLYQHL
31 (SEQ. ID NO:57), RDVHVGIFEAMKMYTVET (SEQ. ID NO:58),
32 GDKLTEIFEAMKIQWTSG (SEQ. ID NO:59), LEGLRAVFESMKMELADE (SEQ.
33 ID NO:60), VADSHDTFAAMKMVWLDL (SEQ. ID NO:61),
34 GLPLQDILES MKIVMTSG (SEQ. ID NO:62), RVPLEAIFEGAKMIWVPNN (SEQ.
35 ID NO:63), PMISHKNFEAMKMLFVPE (SEQ. ID NO:64),
36 KLGLPAMFEAMKMEWHPS (SEQ. ID NO:65), QPSLLSIFEAMKMQASLM (SEQ.
37 ID NO:66), LLELRNFEAMKMEWQIS (SEQ. ID NO:67),
38 DEELNQIFEAMKMYPLVHVTK (SEQ. ID NO:68), SNLVSLLSQKILWTD PQSFG
39 (SEQ. ID NO:70), LFLHDFLNAQKVELYPVTSSG (SEQ. ID NO:71),
40 SDINALLSTQKIYWAH (SEQ. ID NO:72), MASSLRQILDSQKMEWRSNAGGS
41 (SEQ. ID NO:74), MAHSLVPIFDAQKIEWRDPFGGS (SEQ. ID NO:75),
42 MGPDLVNIFEAQKIEWHPLTGGS (SEQ. ID NO:76),
43 MAFSLRSILEAQKMELRNTPGGS (SEQ. ID NO:77),
44 MAGGLNDIFEAQKIEWHEDTGGS (SEQ. ID NO:78),
45 MSSYLAPIFEAQKIEWHSAYGGS (SEQ. ID NO:79),
46 MAKALQKILEAQKMEWRSHPGGS (SEQ. ID NO:80),
47 MAFQLCKIFYAQKMEWHGVGGGS (SEQ. ID NO:81),
48 MAGSLSTIFDAQKIEWHVGKGGGS (SEQ. ID NO:82),
49 MAQQLPDIFDAQKIEWRIAGGGS (SEQ. ID NO:83),
50 MAQRLFHILDAQKIEWHGPKGGGS (SEQ. ID NO:84),
51 MAGCLGPIFEAQKMEWRHFVGGGS (SEQ. ID NO:85),
52 MAWSLKPIFDAQKIEWHSPGGGS (SEQ. ID NO:86),
53 MALGLTRILDAQKIEWHRDSGGS (SEQ. ID NO:87),
54 MAGSLRQILDAQKIEWRRPLGGS (SEQ. ID NO:88), and
55 MADRLAYILEAQKMEWHPHKGGGS (SEQ. ID NO:89).

1 8. A purified preparation of a biotinylated
2 protein, said protein comprising a biotinylation peptide less
3 than 50 amino acids in length.

9. The purified protein of Claim 8, wherein said biotinylation peptide is $LX_1X_2IX_3X_4X_5X_6KX_7X_8X_9X_{10}$ (SEQ. ID NO:1), where

X_1 is any amino acid;

X_2 is any amino acid other than L, V, I, W, F, Y;

X_3 is F or L, X_4 is E or D;

X_5 is A, G, S, or T;

X_6 is Q or M;

X_7 is I, M, or V;

X_8 is E, L, V, Y, or I;

X_9 is W, Y, V, F, L, or I; and

X_{10} is R, H, or any amino acid other than D or E.

10. The purified protein of Claim 8, wherein said biotinylation peptide is selected from the group consisting of LEEVDSTSSAIFDAMKMVWISPTEFR (SEQ. ID NO:14), QGDRDETLPILRAMKMEVYNPGGHEK (SEQ. ID NO:15), SKCSYSHDLKIFEAQKMLVHSLRVMYNY (SEQ. ID NO:16), MASSDDGLLTIFDATKMMFIRT (SEQ. ID NO:17), SYMDRTDVPTILEAMKMELHTTPWACR (SEQ. ID NO:18), SFPPSLPDKNIFEAMKMYVIT (SEQ. ID NO:19), SVVPEPGWDGPFESMKMVYHSGAQSGQ (SEQ. ID NO:20), VRHLPPPLPALFDAMKMEFVTSVQF (SEQ. ID NO:21), DMTMPTGMTKIFEAMKMEVST (SEQ. ID NO:22), ATAGPLHEPDIFLAMKMEVVDVTNKAGQ (SEQ. ID NO:23), SMWETLNAQKTVLL (SEQ. ID NO:24), SHPSQLMTNDIFEGMKMLYH (SEQ. ID NO:25), TSELSKLDATIFAAMKMOWWNPQ (SEQ. ID NO:27), VMETGLDLRPILTGMKMDWIPK (SEQ. ID NO:28), LHHILDAQKMVWNHR (SEQ. ID NO:30), PQGIFEAQKMLWRS (SEQ. ID NO:31), LAGTFEALKMAWHEH (SEQ. ID NO:32), LNAIFEAMKMEYSG (SEQ. ID NO:33), LGGIFEAMKMELRD (SEQ. ID NO:34), LLRTFEAMKMDWRNG (SEQ. ID NO:35), LSTIMEGKMKYIORS (SEQ. ID NO:36), LSDIFEAMKMVYRPC (SEQ. ID NO:37), LESMLEAMKMOWNPQ (SEQ. ID NO:38), LSDIFDAMKMVYRPQ (SEQ. ID NO:39), LAPFFESMKMVWREH (SEQ. ID NO:40), LKGIFEAMKMEYTAM (SEQ. ID NO:41), LEGIFEAMKMEYSNS (SEQ. ID NO:42), LLQTFDAMKMEWLPK (SEQ. ID NO:43), VFDILEAQKVVTTLRF (SEQ. ID NO:44), LVSMFDGMKMEWCTL (SEQ. ID NO:45), LEPIFEAMKMDWRLE (SEQ. ID NO:46), LKEIFEAGMKMEFVKP (SEQ. ID NO:47),

26 LGGILEAQKMLYRGN (SEQ. ID NO:48), RPVLENIFEAMKMEVWKP (SEQ. ID
27 NO:50), RSPIAEIFEAMKMEYRET (SEQ. ID NO:51), QDSIMPIFEAMKMSWHVN
28 (SEQ. ID NO:52), DGVLFPIFESMKMIRLET (SEQ. ID NO:53),
29 VSRMTNFEAMKMIYHDL (SEQ. ID NO:54), DVLLPTVFEAMKMYITK (SEQ. ID
30 NO:55), PNDLERIFDAMKIVTVHS (SEQ. ID NO:56), TRALLEIFDAQKMLYQHL
31 (SEQ. ID NO:57), RDVHVGIFEAMKMYTVET (SEQ. ID NO:58),
32 GDKLTEIFEAMKIQWTSG (SEQ. ID NO:59), LEGLRVFEAMKMELADE (SEQ.
33 ID NO:60), VADSHDTFAAMKMVWLDL (SEQ. ID NO:61),
34 GLPLQDILESMKIVMTSG (SEQ. ID NO:62), RVPLEAIFEGAKMIWVPNN (SEQ.
35 ID NO:63), PMISHKNFEAMKMLFVPE (SEQ. ID NO:64),
36 KGLPAMFEAMKMEWHPS (SEQ. ID NO:65), QPSLLSIFEAMKMQASLM (SEQ.
37 ID NO:66), LLELRNFEAMKMEWQIS (SEQ. ID NO:67),
38 DEELNQIFEAMKMYPLVHVTK (SEQ. ID NO:68), SNLVSLLSQKILWTDPOQSG
39 (SEQ. ID NO:70), LFLHDFLNAQKVELYPVTSSG (SEQ. ID NO:71),
40 SDINALLSTQKIYWAH (SEQ. ID NO:72), MASSLRQILDSQKMEWRSNAGGS
41 (SEQ. ID NO:74), MAHSLVPIFDAQKIEWRDPFGGS (SEQ. ID NO:75),
42 MGPDLVNIFEAQKIEWHPLTGGGS (SEQ. ID NO:76),
43 MAFSLRSILEAQKMELRNTPGGGS (SEQ. ID NO:77),
44 MAGGLNDIFEAQKIEWHEDTGGGS (SEQ. ID NO:78),
45 MSSYLAPIFEAQKIEWHSAYGGGS (SEQ. ID NO:79),
46 MAKALQKILEAQKMEWRSHPGGS (SEQ. ID NO:80),
47 MAFQLCKIFYAQKMEWHGVGGGS (SEQ. ID NO:81),
48 MAGSLSTIFDAQKIEWHVGKGGGS (SEQ. ID NO:82),
49 MAQQLPDIFDAQKIEWRIAGGGGS (SEQ. ID NO:83),
50 MAQRLFHILDAQKIEWHGPKGGGS (SEQ. ID NO:84),
51 MAGCLGPIFEAQKMEWRHFVGGGS (SEQ. ID NO:85),
52 MAWSLKPIFDAQKIEWHSPGGGS (SEQ. ID NO:86),
53 MALGLTRILDAQKIEWHRDSGGGS (SEQ. ID NO:87),
54 MAGSLRQILDAQKIEWRRPLGGGS (SEQ. ID NO:88), and
55 MADRLAYILEAQKMEWHPHKGGGS (SEQ. ID NO:89).

1 11. A kit for biotinylating a protein, the kit
2 comprising a recombinant DNA expression polynucleotide that
3 encodes a biotinylation peptide less than 50 amino acids which
4 can be fused in frame with a protein by inserting the coding
5 sequence for the protein.

12. The kit of Claim 11, wherein said biotinylation peptide is $LX_1X_2IX_3X_4X_5X_6KX_7X_8X_9X_{10}$ (SEQ. ID NO:1), where

X_1 is any amino acid;

X_2 is any amino acid other than L, V, I, W, F, Y;

X_3 is F or L, X_4 is E or D;

X_5 is A, G, S, or T;

X_6 is Q or M;

X_7 is I, M, or V;

X_8 is E, L, V, Y, or I;

X_9 is W, Y, V, F, L, or I; and

X_{10} is R, H, or any amino acid other than D or E.

13. The kit of Claim 11, wherein said biotinylation peptide is selected from the group consisting of

LEEVDSTSSAIFDAMKMWISPTEFR (SEQ. ID NO:14),

QGDRDETLPILRAMKMEVYNPGGHEK (SEQ. ID NO:15),

SKCSYSHDLKIFEAQKMLVHSLRVMYNY (SEQ. ID NO:16),

MASSDDGLLTIFDATKMMFIRT (SEQ. ID NO:17),

SYMDRTDVPTILEAMKMEHTTPWACR (SEQ. ID NO:18),

SFPPSLPDKNIFEAMKMYVIT (SEQ. ID NO:19),

SVVPEPGWDGPFESMKMVYHSGAQSGQ (SEQ. ID NO:20),

VRHLPPPLPALFDAMKMEFVTSVQF (SEQ. ID NO:21),

DMTMPTGMTKIFEAMKMEVST (SEQ. ID NO:22),

ATAGPLHEPDIFLAMKMEVVDVTNKAGQ (SEQ. ID NO:23), SMWETLNAQKTVLL (SEQ. ID NO:24),

SHPSQLMTNDIFEGMKMLYH (SEQ. ID NO:25),

TSELSKLDATIFAAMKMOWWNP (SEQ. ID NO:27),

VMETGLDLRPILTGMKMDWIPK (SEQ. ID NO:28), LHHILDAQKMWVWVHR (SEQ. ID NO:30),

PQGIFEAQKMLWRS (SEQ. ID NO:31), LAGTFEALKMAWHEH (SEQ. ID NO:32),

LNAIFEAMKMEYSG (SEQ. ID NO:33), LGGIFEAMKMELRD (SEQ. ID NO:34),

LLRTFEAMKMDWRNG (SEQ. ID NO:35), LSTIMEGMKMYIQRS (SEQ. ID NO:36),

LSDIFEAMKMYVRPC (SEQ. ID NO:37), LESMLEAMKMOWNPQ (SEQ. ID NO:38),

LSDIFDAMKMYVRPQ (SEQ. ID NO:39), LAPFFESMKMVWREH (SEQ. ID NO:40),

LKGIFEAMKMEYTAM (SEQ. ID NO:41), LEGIFEAMKMEYSNS (SEQ. ID NO:42),

LLQTFDAMKMEWLPK (SEQ. ID NO:43), VFDILEAQKVVTILRF (SEQ. ID NO:44),

LVSMTFDGMKMEWCTL (SEQ. ID NO:45), LEPIFEAMKMDWRLE (SEQ. ID NO:46),

LKEIFEGMKMEFVKP (SEQ. ID NO:47), LGGILEAQKMLYRGN (SEQ. ID NO:48),

RPVLENIFEAMKMEVWKP (SEQ. ID NO:49).

27 NO:50), RSPIAEIFEAMKMEYRET (SEQ. ID NO:51), QDSIMPIFEAMKMSWHVN
28 (SEQ. ID NO:52), DGVLFPIFESMKMIRLET (SEQ. ID NO:53),
29 VSRTMTNFEAMKMIYHDL (SEQ. ID NO:54), DVLLPTVFEAMKMYITK (SEQ. ID
30 NO:55), PNDLERIFDAMKIVTVHS (SEQ. ID NO:56), TRALLEIFDAQKMLYQHL
31 (SEQ. ID NO:57), RDVHVGIFEAMKMYTVET (SEQ. ID NO:58),
32 GDKLTEIFEAMKIQWTSG (SEQ. ID NO:59), LEGLRAVFESMKMELADE (SEQ.
33 ID NO:60), VADSHDTFAAMKMVWLDL (SEQ. ID NO:61),
34 GLPLQDILESMKIVMTSG (SEQ. ID NO:62), RVPLEAIFEGAKMIWVPNN (SEQ.
35 ID NO:63), PMISHKNFEAMKMLFVPE (SEQ. ID NO:64),
36 KLGLPAMFEAMKMEWHPS (SEQ. ID NO:65), QPSLLSIFEAMKMQASLM (SEQ.
37 ID NO:66), LLELRNFEAMKMEWQIS (SEQ. ID NO:67),
38 DEELNQIFEAMKMYPLVHVTK (SEQ. ID NO:68), SNLVSLLSQKILWTD PQSFG
39 (SEQ. ID NO:70), LFLHDFLNAQKVELYPVTSSG (SEQ. ID NO:71),
40 SDINALLSTQKIYWAH (SEQ. ID NO:72), MASSLRQILDSQKMEWRSNAGGS
41 (SEQ. ID NO:74), MAHSLVPIFDAQKIEWRDPFGGS (SEQ. ID NO:75),
42 MGPDLVNIFEAQKIEWHPLTGGS (SEQ. ID NO:76),
43 MAFSLRSILEAQKMELRNTPGGS (SEQ. ID NO:77),
44 MAGGLNDIFEAQKIEWHEDTGGS (SEQ. ID NO:78),
45 MSSYLAPIFEAQKIEWHSAYGGS (SEQ. ID NO:79),
46 MAKALQKILEAQKMEWRSHPGGS (SEQ. ID NO:80),
47 MAFQLCKIFYAQKMEWHGVGGGS (SEQ. ID NO:81),
48 MAGSLSTIFDAQKIEWHVGKGGGS (SEQ. ID NO:82),
49 MAQQLPDIFDAQKIEWRIAGGGS (SEQ. ID NO:83),
50 MAQRLFHILDAQKIEWHGPKGGGS (SEQ. ID NO:84),
51 MAGCLGPIFEAQKMEWRHFVGGGS (SEQ. ID NO:85),
52 MAWSLKPIFDAQKIEWHSPGGGS (SEQ. ID NO:86),
53 MALGLTRILDAQKIEWHRDSGGGS (SEQ. ID NO:87),
54 MAGSLRQILDAQKIEWRRPLGGGS (SEQ. ID NO:88), and
55 MADRLAYILEAQKMEWHPHKGGGS (SEQ. ID NO:89).

1 14. The kit of Claim 11 wherein the expression
2 polynucleotide is transformed into a host cell.

1 15. The kit of Claim 12 wherein the host cell
2 overexpresses a biotin protein ligase.

1 16. A method for identifying a biotinylating
2 enzyme, said method comprising:

(a) on a surface of a substrate, providing a fusion protein comprising a protein and a biotinylation peptide of less than 50 amino acids in length;
(b) in a predefined region of the surface, contacting said fusion protein with an enzyme; and
(c) determining whether the fusion protein has been biotinylated.

17. The method of Claim 16, wherein said fusion protein is contacted with a plurality of different enzymes, wherein each of the different enzymes is in a different predefined region.

18. The method of Claim 16, wherein the step of determining whether the fusion protein has been biotinylated further comprises the steps of;

- i) treating the fusion protein with labeled streptavidin, wherein the labeled streptavidin binds to biotinylated fusion protein;
- ii) washing the substrate to remove unbound labeled streptavidin; and
- iii) detecting the presence of labeled streptavidin which has bound to biotinylated fusion protein.

19. The method of Claim 16, wherein the substrate comprises a support having a plurality of wells.

20. The method of Claim 19, wherein the support having a plurality of wells is a 96-well microtiter plate.

21. The method of Claim 16, wherein said biotinylation peptide is $LX_1X_2IX_3X_4X_5X_6KX_7X_8X_9X_{10}$ (SEQ. ID NO:1), where

- X_1 is any amino acid;
- X_2 is any amino acid other than L, V, I, W, F, Y;
- X_3 is F or L, X_4 is E or D;
- X_5 is A, G, S, or T;
- X_6 is Q or M;

- 9 X₇ is I, M, or V;
10 X₈ is E, L, V, Y, or I;
11 X₉ is W, Y, V, F, L, or I; and
12 X₁₀ is R, H, or any amino acid other than D or E.

1 22. The method of Claim 16, wherein said
2 biotinylation peptide is selected from the group consisting of
3 LEEVDSTSSAIFDAMKMVWISPTEFR (SEQ. ID NO:14),
4 QGDRDETLPMILRAMKMEVYNPGGHEK (SEQ. ID NO:15),
5 SKCSYSHDLKIFEAQKMLVHSYLRVMYNY (SEQ. ID NO:16),
6 MASSDDGLLTIFDATKMMFIRT (SEQ. ID NO:17),
7 SYMDRTDVPTILEAMKMEHTTPWACR (SEQ. ID NO:18),
8 SFPPSLPDKNIFEAMKMYVIT (SEQ. ID NO:19),
9 SVVPEPGWDGPFESMKMVYHSGAQSGQ (SEQ. ID NO:20),
10 VRHLPPPLPALFDAMKMEFVTSVQF (SEQ. ID NO:21),
11 DMTMTGTMKIFEAMKMEVST (SEQ. ID NO:22),
12 ATAGPLHEPDIFLAMKMEVVDVTNKAGQ (SEQ. ID NO:23), SMWETLNAQKTVLL
13 (SEQ. ID NO:24), SHPSQLMTNDIFEGMKMLYH (SEQ. ID NO:25),
14 TSELSKLDATIFAAMKMOWWNP (SEQ. ID NO:27),
15 VMETGLDLRPILTGMKMDWIPK (SEQ. ID NO:28), LHHILDAQKMVWNHR (SEQ.
16 ID NO:30), PQGIFEAQKMLWRS (SEQ. ID NO:31), LAGTFEALKMAWHEH
17 (SEQ. ID NO:32), LNAIFEAMKMEYSG (SEQ. ID NO:33),
18 LGGIFEAMKMELRD (SEQ. ID NO:34), LLRTFEAMKMDWRNG (SEQ. ID
19 NO:35), LSTIMEGMKMYIQRS (SEQ. ID NO:36), LSDIFEAMKMVYRPC (SEQ.
20 ID NO:37), LESMLEAMKMOWNPQ (SEQ. ID NO:38), LSDIFDAMKMVYRPQ
21 (SEQ. ID NO:39), LAPFFESMKMVWREH (SEQ. ID NO:40),
22 LKGIFEAMKMEYTAM (SEQ. ID NO:41), LEGIFEAMKMEYSNS (SEQ. ID
23 NO:42), LLQTFDAMKMEWLPK (SEQ. ID NO:43), VFDILEAQKVVTLR (SEQ.
24 ID NO:44), LVSMFDGMKMEWKT (SEQ. ID NO:45), LEPIFEAMKMDWRLE
25 (SEQ. ID NO:46), LKEIFEGMKMEFVKP (SEQ. ID NO:47),
26 LGGILEAQKMLYRGN (SEQ. ID NO:48), RPVLENIFEAMKMEVWKP (SEQ. ID
27 NO:50), RSPIAEIFEAMKMEYRET (SEQ. ID NO:51), QDSIMPIFEAMKMSWHVN
28 (SEQ. ID NO:52), DGVLPPIFESMKMIRLET (SEQ. ID NO:53),
29 VSRTMTNFEAMKMIYHDL (SEQ. ID NO:54), DVLLPTVFEAMKMYITK (SEQ. ID
30 NO:55), PNDLERIFDAMKIVTVHS (SEQ. ID NO:56), TRALLEIFDAQKMLYQHL
31 (SEQ. ID NO:57), RDVHVGIFEAMKMYTVET (SEQ. ID NO:58),
32 GDKLTEIFEAMKIQWTSG (SEQ. ID NO:59), LEGLRVFEAMKMELADE (SEQ.
33 ID NO:60), VADSHDTFAAMKMVWLD (SEQ. ID NO:61),

34 GLPLQDILESMKIVMTSG (SEQ. ID NO:62), RVPLEAIFEGAKMIWVPNN (SEQ.
35 ID NO:63), PMISHKNFEAMKMLFVPE (SEQ. ID NO:64),
36 KLGLPAMFEAMKMEWHPS (SEQ. ID NO:65), QPSLLSIFEAMKMQASLM (SEQ.
37 ID NO:66), LLELRSNFEAMKMEWQIS (SEQ. ID NO:67),
38 DEELNQIFEAMKMYPLVHVTK (SEQ. ID NO:68), SNLVSLLSQKILWTDPOSG
39 (SEQ. ID NO:70), LFLHDFLNAQKVELYPVTSSG (SEQ. ID NO:71),
40 SDINALLSTQKIYWAH (SEQ. ID NO:72), MASSLRQILDSQKMEWRSNAGGS
41 (SEQ. ID NO:74), MAHSLVPIFDAQKIEWRDPFGGS (SEQ. ID NO:75),
42 MGPDLVNIFEAQKIEWHPLTGGS (SEQ. ID NO:76),
43 MAFSLRSILEAQKMELRNTPGGS (SEQ. ID NO:77),
44 MAGGLNDIFEAQKIEWHEDTGGS (SEQ. ID NO:78),
45 MSSYLAPIFEAQKIEWHSAYGGS (SEQ. ID NO:79),
46 MAKALQKILEAQKMEWRSHPGGS (SEQ. ID NO:80),
47 MAFQLCKIFYAQKMEWHGVGGGS (SEQ. ID NO:81),
48 MAGSLSTIFDAQKIEWHVGKGGGS (SEQ. ID NO:82),
49 MAQQLPDIFDAQKIEWRIAGGGS (SEQ. ID NO:83),
50 MAQRLFHILDAQKIEWHGPKGGGS (SEQ. ID NO:84),
51 MAGCLGPIFEAQKMEWRHFGVGGGS (SEQ. ID NO:85),
52 MAWSLKPIFDAQKIEWHSPGGGS (SEQ. ID NO:86),
53 MALGLTRILDAQKIEWHRDSGGGS (SEQ. ID NO:87),
54 MAGSLRQILDAQKIEWRRPLGGGS (SEQ. ID NO:88), and
55 MADRLAYILEAQKMEWHPHKGGGS (SEQ. ID NO:89).

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US94/08528

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) : C07H 21/04; C12N 15/11, 15/63, 15/65
US CL : 435/69.1, 69.7, 172.3, 320.1; 436/56, 86; 530/324, 325; 536/23.4
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/69.1, 69.7, 172.3, 320.1; 436/56, 86; 530/324, 325; 536/23.4

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, BIOSIS, MEDLINE, EMBASE
search terms: biotin, protein, fusion protein, express, plasmid

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
<u>X</u> Y	Journal of Biological Chemistry, Volume 259, No. 20, issued 25 October 1984, S.O. Freytag et al., "Molecular Cloning of a cDNA for Human Pyruvate Carboxylase", pages 12831-12837, see entire document.	<u>1,2,4,5,8</u> 11,14-18
<u>X</u> Y	Proceedings of the National Academy of Science, Volume 82, issued September 1985, V.L. Murtif et al., "Cloning and Expression of the 1.3S Biotin-containing Subunit of Transcarboxylase", pages 5617-5621, see entire document.	<u>1,2,4,5,8</u> 11,14-18

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
A document defining the general state of the art which is not considered to be of particular relevance	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
E earlier document published on or after the international filing date	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*-&* document member of the same patent family
O document referring to an oral disclosure, use, exhibition or other means	
P document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

07 OCTOBER 1994

Date of mailing of the international search report

07 NOV 1994

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

DIAN C. JACOBSON

Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US94/08528

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
<u>X</u> Y	Archives of Biochemistry and Biophysics, Volume 254, No. 2, issued 01 May 1987, A.-M. Lamhonwah et al., "Sequence Homology around the Biotin-Binding Site of Human Propionyl-CoA Carboxylase and Pyruvate Carboxylase", pages 631-636, see especially abstract.	<u>1,2,4,5,8</u> 11,14-18
<u>X</u> Y	Archives of Biochemistry and Biophysics, Volume 266, No. 1, issued October 1988, K.G. Thampy et al., "A Rapid Purification Method for Rat Liver Pyruvate Carboxylase and Amino Acid Sequence Analyses of NH ₂ -Terminal and Biotin Peptide", pages 270-276, see entire document.	<u>1,2,4,5,8</u> 11,14-18
A	Journal of Biological Chemistry, Volume 266, No. 18, issued 25 June 1991, K.E. Reed et al., "Escherichia coli Exports Previously Folded and Biotinated Protein Domains", pages 11425-11428, see entire document.	1-22